Genetic Basis of Adaptation: SNP-level Detection Method, Genetic Architecture of Adaptative Traits in *Drosophila melanogaster*, and Population Genomics of Saline and Freshwater Populations of *Eurytemora affinis*

By

Tiago da Silva Ribeiro

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Integrative Biology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 01/16/2024

The dissertation is approved by the following members of the Final Oral Committee:
John E. Pool, Professor, Genetics and Integrative Biology
Cécile M. Ané, Professor, Botany and Statistics
Guilherme J. M. Rosa, Professor, Animal & Dairy Sciences
Sean Schoville, Associate Professor, Entomology
David A. Wassarman, Professor, Genetics

Acknowledgements

I had the privilege of meeting and working alongside incredible people during my time here, without whom the work presented in this thesis would not have been possible. I thank John Pool, my advisor, for welcoming me into his laboratory and supporting my every step. John taught me science and biology, but also how to keep going.

I would like to thank the members of my doctoral committee, Cécile Ané, Guilherme Rosa, Sean Schoville, and David Wassarman, for their guidance, encouragement, and support.

To the members of the Pool Lab. Yuheng Huang, for teaching me how to work with live flies in the laboratory. Matthew Lollar and Quentin Sprengelmeyer, for the help with bench work. Jamie Freeman, for the massive amount of DNA extractions and library prep. Chris, Jeremy, Siyuan, Maud, Max, and Myron, for the questions, for the answers, and for keeping science exciting.

To the evolution community and the Integrative Biology graduate students, for the camaraderie. Megan Frayer, Jered Stratton, Joseph Sardina, and Linh Nguyen, for organizing the Evolution Seminar Series with me, as well as Martin Bontrager and Juanita Diaz for helping me navigate graduate school in my first years.

To the staff of the Laboratory of Genetics and the Integrative Biology Department, for the assistance with piles of paperwork. Nazan Gillie and Kayla Pelland, for being there and showing me a way forward in the hardest times. I

would like to thank Gale Oakes, for believing in me and making teaching a walk in the park—sometimes literally!

To the NSF, NIH, and UW-Madison, for the funding and infrastructure.

To the Madison Community Cooperative, for housing me for over four years, enriching my life outside academia, and drastically changing the way I see the world. Particularly Friends and Ambrosia Housing Cooperatives, and every single overly passionate human being that shared a roof and a meal with me.

To my friends in Brazil, who kept in touch with me on a daily—if not hourly—basis and soothed the impact of distance.

To my family, for the love and waiting for me.

Table of Contents

Abstract

Understanding the genetic basis of adaptation is a central goal of evolutionary biology. Herein, I addressed methodological and empirical questions using computer simulations and biological empirical. In my second chapter, I assessed the power of a SNP-level statistic to detect genomic signatures of selection and compared it to window-based metrics. I found that the different approaches have complementary power to detect distinct kinds of selective events.

In the third and fourth chapters, I investigated the genetic basis of adaptive traits from natural populations of *Drosophila melanogaster*. In chapter three, I leveraged two newly generated panels of recombinant inbred lines and showed that adaptive traits often have loci of detectable effect sizes. I showed that there is no evidence of strong gene-by-gene interaction involving adaptive loci, but gene-by-environment interactions affected the effect size of loci underlying pigmentation. These results imply that if epistasis plays a role in adaptation, it is not through large changes in phenotypic effects. Chapter four focused on adaptive pigmentation. I studied three pigmentation traits in mapping crosses generated with strains from the same population pair. No quantitative trait locus (QTL) was shared across all the trait mappings. The three most common QTLs had the strongest effect sizes and overlapped the pigmentation-related genes *ebony*, *tan*, and *yellow*. QTLs had a higher

likelihood of being shared between two trait mappings if they shared at least one parental strain, suggesting that several loci of moderate to strong effect underly adaptive pigmentation, likely through selection on standing variation with partial allele frequency changes.

Lastly, I used four populations of *Eurytemora affinis* from across salinity and temporal gradients to search for biological functions under selection. A brackish and a saltwater population that colonized the Saint Lawrence estuary approximately 17 kya, and two invasive freshwater populations that colonized the Great Lakes in the last ~70 years. I found that ion transport was an important biological function under selection across salinities and timescales, and ion transport regulation was uniquely enriched in the older timescales.

Overall, I contributed methodological and empirical advances to the field, showing that detectable adaptive variants can be common in nature.

Chapter 1: Introduction

Evolutionary biology studies the mechanisms underlying the origin and maintenance of biodiversity. A central question in the field is how organisms adapt to their environment. Adaptations are the result of natural selection, an evolutionary mechanism that requires heritable variation among individuals that has an effect on fitness (survival and reproduction). Individuals with traits promoting higher fitness will have more offspring over their lifetime and these heritable traits will be passed on to their offspring. Over time the beneficial trait will become more prevalent in the population, which in turn becomes better adapted to its environment. In a rapidly changing world, with virtually no habitat unmodified by human activity, understanding how populations adapt to novel environments is ever more pressing.

Important aspects regarding the genetic basis of adaptation are still unresolved. An allele underlying a beneficial trait is expected to increase in frequency in the population as the trait becomes more frequent. When a beneficial allele is selected and increases in frequency until it is fixed in a population, it reduces the genetic diversity of the neutral genetic diversity linked to it, in a process called selective sweep (Smith & Haigh 1974). There are still debates regarding (1) the origin of the beneficial allele (new mutations *versus* standing variation) and (2) how often the beneficial allele actually reaches fixation in the population. Although we know that the

answer to these questions varies case by case, depending on the trait, species, or even population being studied, their relative contributions and when each is more likely to take place is still unknown. Selection on standing variation is thought to be a crucial mechanism for rapid adaptation (Barrett & Schluter 2008).

Both factors regarding the origin and fate of adaptive alleles have implications on how the selective sweep impacts nearby genetic diversity. When the beneficial allele is found in a single haplotype, it creates a hard sweep. This has a stronger effect on the linked genetic diversity than when the beneficial allele is initially in multiple haplotypes in the population, causing two or more of these haplotypes to increase in frequency, creating a soft sweep (Hermisson & Pennings 2017). Hard and soft sweeps, as well as complete and partial sweeps will affect the neutral genetic diversity linked to the beneficial mutation differently, changing the signatures of the action of natural selection in the genome (Figure 1). That said, although pervasive evidence of soft sweeps has been shown in humans (Schrider & Kern 2017) and *Drosophila melanogaster* (Garud *et al*. 2015), the models underlying these results have been hotly debated (Harris *et al*. 2018, Schrider & Kern 2018).

The number of loci most commonly underlying adaptive changes is also a topic of debate. Polygenic adaptation posits that small frequency changes in many loci underlying complex traits could produce adaptive changes without detectable selective sweeps (Pritchard *et al*. 2010, Pritchard

& Di Rienzo 2010). However, empirical studies have also shown examples of polygenic adaptation with loci of detectable effect size reaching intermediate frequencies (Barghi *et al*. 2019, Höllinger *et al*. 2023) and adaptation due to a single major locus of strong effect (Bersaglieri *et al*. 2004, Hoekstra *et al*. 2006).

Which loci are involved in adaptation and their phenotypic effect, however, seems to vary depending on the scale of the study or even which populations are studied (Huber *et al*. 2015), and in some cases it will vary among individuals of the same population as well (Bastide *et al*. 2016). And given that selection acts on the phenotypic effects, other factors that affect phenotypic effects of beneficial variants, such as gene-by-gene and gene-by-environment interactions, could also affect natural selection. Gene-by-environment interactions have been shown to alter the effect of adaptive alleles in *D. melanogaster* and wheat, for example (Fry *et al*. 1998, Mathews *et al*. 2008). But the role of gene-by-gene interactions, also known as epistasis, has been debated for nearly a century, with some defending that it is irrelevant to adaptation (Fisher 1930, Hill *et al*. 2008, Crow 2010) and others defending the contrary (Wright 1931, Hansen 2013). There are many aspects underlying the genotype-phenotype map, and the understanding the degree to which they affect the adaptive process is crucial to comprehend how organisms adapt to their environments.

In this thesis, I've aimed to address questions regarding the genetic basis of adaptation by (1) testing the power of a method to detect different kinds of selective sweeps, (2) investigating the genetic architecture of an adaptive trait, and (3) scanning the genome of an invasive population for signatures of selection.

In the second chapter, I define an approach to scan the genome for signatures of selection based on SNP-level F_{ST}. F_{ST} is a statistic commonly used to measure differentiation between two or more populations. The differentiation between two populations is expected to increase in the genomic regions targeted by natural selection, and so elevated F_{ST} is a signature of selective sweeps (Lewontin & Krakauer 1973). Different kinds of selective sweeps might leave different signatures, as soft sweeps, in particular, might have a narrower effect on the genome than hard sweeps. Therefore, I tested an approach focusing on the highest F_{ST} value of an individual SNP versus more traditional approaches using window-based genetic patterns. I compared their power detecting different kinds of sweeps using computer simulations and models based on human and *D. melanogaster* parameters and used empirical *D. melanogaster* data as a proof of concept.

In the third and fourth chapters, I focused on the genetic basis of adaptive traits from natural populations of *D. melanogaster*. In the third chapter, I used recombinant inbred lines to map the loci underlying a

series of adaptive traits, identify their effect sizes, and subsequently examine the effects of gene-by-gene and gene-by-environment interactions on the adaptive loci. Genetic and environmental interactions are known to affect the phenotypic effect of loci underlying quantitative traits, and my aim in this chapter was to understand how they could affect adaptive traits. In the fourth chapter, I focused on the genetic architecture of adaptive pigmentation. A previous study had shown that adaptive pigmentation has a variable genetic basis (Bastide *et al*. 2016), so I expanded the experimental design of that study and focused on an Ethiopian population with the darkest known flies (Bastide *et al*. 2014). My main goal was to investigate the number of loci involved in this adaptation, and how often the adaptive loci were needed to produce darker flies. I used seven dark inbred strains from the adapted population and three light inbred strains from a population within the species ancestral range and generated twenty-one mapping crosses. Pigmentation shows thermal plasticity in flies (David *et al*. 1990), so to understand the temperature effect on adaptive loci I also studied a subset of the crosses in a colder environment, more similar to the Ethiopian population.

In the fifth chapter, I searched for candidate genes under selection across a salinity gradient and at two different timescales by comparing different populations in the Atlantic clade of the *Eurytemora affinis* complex copepods. Unlike the previous two chapters, here I used an approach

starting from the genomic data to find the biological functions underlying local adaptation. My main goal was to understand the genetic basis of adaptive change across environmental conditions and temporal scales. I compared two freshwater populations that invaded the Great Lakes of North America within the last ~70 years and two populations that colonized the Saint Lawrence estuary after the Last Glacial Maximum ~17 kya (Lee 2000), with different salinity conditions: a brackish and a saltwater population.

Overall, I used a diverse set of approaches to further our understanding of the genetic basis of the adaptive process. I contributed methodological approaches to the study of selective sweeps and empirical results on the genetic basis of adaptive evolution.

References

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biology*, 17(2), e3000128. https://doi.org/10.1371/journal.pbio.3000128.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38-44. https://doi.org/10.1016/j.tree.2007.09.008

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., & Pool, J. E. (2016). A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics*, 204(3), 1307–1319. https://doi.org/10.1534/genetics.116.192492.

Bastide, H., Yassin, A., Johanning, E. J., & Pool, J. E. (2014). Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan

Africa. *BMC Evolutionary Biology*, 14(1), 179. https://doi.org/10.1186/s12862-014-0179-y.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6), 1111-1120. https://doi.org/10.1086/421051.

Crow, J. F. (2010). On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1241–1244. https://doi.org/10.1098/rstb.2009.0275.

David, J. R., Capy, P., & Gauthier, J.-P. (1990). Abdominal pigmentation and growth temperature in *Drosophila melanogaster*: similarities and differences in the norms of reaction of successive segments. *Journal of Evolutionary Biology*, 3(5–6), 429–445. https://doi.org/10.1046/j.1420-9101.1990.3050429.x.

Fisher, R. A. (1930). The genetical theory of natural selection (pp. xiv, 272). Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Fry, J. D., Nuzhdin, S. V., Pasyukova, E. G., & Mackay, T. F. (1998). QTL mapping of genotype–environment interaction for fitness in *Drosophila melanogaster*. *Genetics Research*, 71(2), 133-141. https://doi.org/10.1017/S0016672398003176.

Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genetics*, 11(2), e1005004. https://doi.org/10.1371/journal.pgen.1005004.

Harris, R. B., Sackman, A., & Jensen, J. D. (2018). On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. *PLoS Genetics*, 14(12), e1007859. https://doi.org/10.1371/journal.pgen.1007859.

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, 8(6), 700-716. https://doi.org/10.1111/2041-210X.12808.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313(5783), 101-104. https://doi.org/10.1126/science.1126121.

Höllinger, I., Wölfl, B., & Hermisson, J. (2023). A theory of oligogenic adaptation of a quantitative trait. *Genetics*, 225(2), iyad139. https://doi.org/10.1093/genetics/iyad139.

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W. O., & Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, 114(5), 515-524. https://doi.org/10.1038/hdy.2015.22

Lee, C. E. (2000). Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". *Evolution*, 54(6), 2014-2027. https://doi.org/10.1111/j.0014-3820.2000.tb01245.x.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74(1), 175-195. https://doi.org/10.1093/genetics/74.1.175.

Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution*, *67*(12), 3501–3511. https://doi.org/10.1111/evo.12214.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genetics*, *4*(2), e1000008. https://doi.org/10.1371/journal.pgen.1000008.

Mathews, K. L., Malosetti, M., Chapman, S., McIntyre, L., Reynolds, M., Shorter, R., & Van Eeuwijk, F. (2008). Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theoretical and Applied Genetics*, 117, 1077-1091. https://doi.org/10.1007/s00122-008-0846-8.

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation–not by sweeps alone. *Nature Reviews Genetics*, 11(10), 665-667. https://doi.org/10.1038/nrg2880.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4), R208-R215. https://doi.org/10.1016/j.cub.2009.11.055.

Schrider, D. R., & Kern, A. D. (2017). Soft sweeps are the dominant mode of adaptation in the human genome. *Molecular Biology and Evolution*, 34(8), 1863-1877. https://doi.org/10.1093/molbev/msx154.

Schrider, D. R., & Kern, A. D. (2018). On the well-founded enthusiasm for soft sweeps in humans: a reply to Harris, Sackman, and Jensen. https://zenodo.org/doi/10.5281/zenodo.1473781.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35. https://doi.org/10.1017/S0016672300014634.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093%2Fgenetics%2F16.2.97.

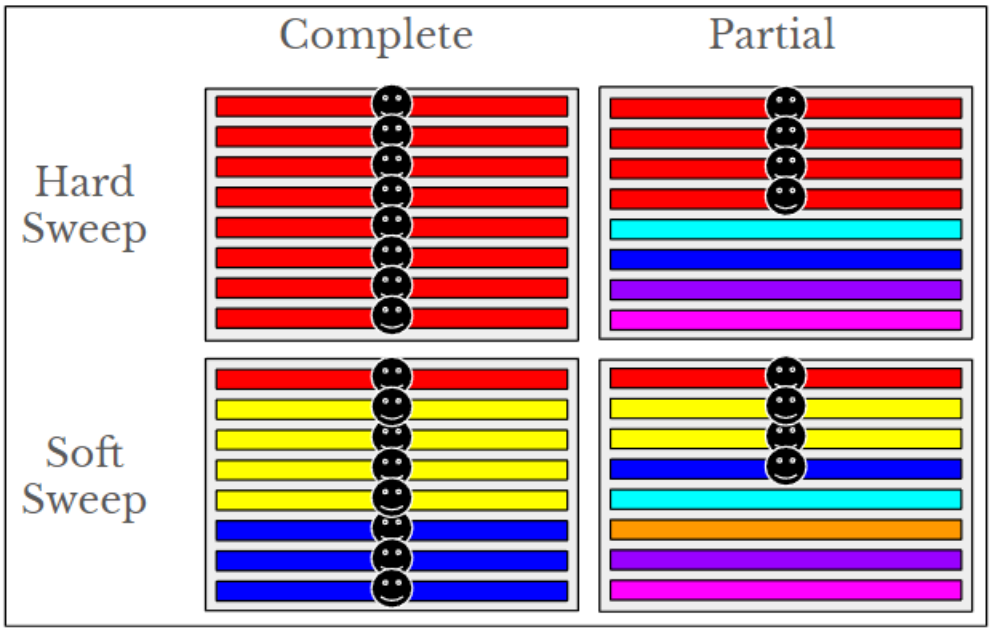


Figure 1. Outcome of different kinds of selective sweeps. Each bar represents an individual and each color represents a different haplotype. The black smiley face represents a beneficial allele that increased in frequency in a population. In hard sweeps, the beneficial mutation is found in a single haplotype, altering more of the region genetic diversity than soft sweeps.

Chapter 2: Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation

A version of this chapter has been published in *Genome Biology and Evolution*.

Citation: da Silva Ribeiro, T., Galván, J. A., & Pool, J. E. (2022). Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation. *Genome Biology and Evolution*, 14(10), evac143. https://doi.org/10.1093/gbe/evac143.

Co-author contribution acknowledgment

We would like to thank José A. Galván contribution to the implementation of the code and demographic models used in this work.

Abstract

Local adaptation can lead to elevated genetic differentiation at the targeted genetic variant and nearby sites. Selective sweeps come in different forms, and depending on the initial and final frequencies of a favored variant, very different patterns of genetic variation may be produced. If local selection favors an existing variant that had already recombined onto multiple genetic backgrounds, then the width of elevated

genetic differentiation (high F_{ST}) may be too narrow to detect using a typical windowed genome scan, even if the targeted variant becomes highly differentiated. We therefore used a simulation approach to investigate the power of SNP-level F_{ST} (specifically, the maximum SNP F_{ST} value within a window, or F_{ST_MaxSNP}) to detect diverse scenarios of local adaptation, and compared it against whole-window F_{ST} and the Comparative Haplotype Identity statistic. We found that F_{ST_MaxSNP} had superior power to detect complete or mostly complete soft sweeps, but lesser power than full-window statistics to detect partial hard sweeps. Nonetheless, the power of F_{ST_MaxSNP} depended highly on sample size, and confident outliers depend on robust precautions and quality control. To investigate the relative enrichment of F_{ST_MaxSNP} outliers from real data, we applied the two F_{ST} statistics to a panel of *Drosophila melanogaster* populations. We found that F_{ST_MaxSNP} had a genome-wide enrichment of outliers compared to demographic expectations, and though it yielded a lesser enrichment than window F_{ST}, it detected mostly unique outlier genes and functional categories. Our results suggest that F_{ST_MaxSNP} is highly complementary to typical window-based approaches for detecting local adaptation, and merits inclusion in future genome scans and methodologies.

Introduction

Geographically distinct populations are exposed to different selective pressures, which may result in local adaptation. The detection of genomic regions under positive selection specific to one population is essential to uncovering the genetic basis of locally adaptive trait variation. Local adaptation can exist between populations with low genome-wide genetic differentiation, and comparing genetic variation between these closely-related populations can allow for much more powerful detection of positive selection than is possible from a single population. In light of that advantage, as well as the potential applicability of genetic mapping and functional approaches to locally adaptive traits, local adaptation has played a key role in our increasing understanding of adaptive evolution at the genetic level (Kawecki & Ebert 2004, Yeaman 2015, Tigano & Friesen 2016). In addition to its importance for evolutionary biology and ecology, the identification of regions under selection has implications for applied fields such as health sciences and agriculture because it can also pinpoint regions of the genome that hold functional diversity (Bamshad & Wooding 2003, Ross-Ibarra *et al*. 2007). There has also been increasing recognition of the importance of local adaptation for a species' future adaptive potential, with implications for conservation genetics and adaptation to climate change (Funk *et al.* 2012, Aitken & Whitlock 2013, Fitzpatrick & Keller 2015).

Population genomic scans for local adaptation compare genetic variation between two or more populations, often searching for specific genomic windows that depart from genome-wide patterns of differentiation in a manner consistent with population-specific natural selection. Positive selection has traditionally been conceptualized and modeled as a selective sweep, which traditionally involves a new beneficial mutation rising to fixation, with strong effects on genetic variation at linked sites (Smith & Haigh 1974; Kaplan *et al*. 1989). However, there are different kinds of selective sweeps, depending on the initial and final frequencies of the favored variant, and different statistical tests for deviations from neutrality vary in their power to detect them.

First, selective sweeps can be classified as hard or soft sweeps. In a hard sweep, only a single original haplotype carrying the advantageous allele is boosted by natural selection. This situation might be expected if selection favors either a newly occurring mutation or else a variant at low enough frequency that only one copy contributes to the sweep by chance. In a soft sweep, two or more distinct haplotypes carrying the beneficial variant increase in frequency. In some cases, soft sweeps occur because the advantageous allele was present in the population, segregating neutrally, prior to the onset of selection (Hermisson & Pennings 2005). But they can also be the result of recurrent mutations or influx of new alleles through migration (Pennings & Hermisson 2006a, 2006b).

Selective sweeps can also be classified as complete or partial sweeps. In a complete sweep, the advantageous allele has reached fixation in the population. In a partial sweep, the advantageous allele is at an intermediary frequency. This may occur either because the sweep is still ongoing, because positive selection ended prior to fixation, or (in the context of local adaptation) because migration continues to supply the non-favored variant. Situations in which a sweep might terminate prematurely include an environmental change, a polygenic trait reaching its new optimum or threshold value, or an allele reaching a balanced equilibrium in a scenario such as heterozygote advantage.

Different kinds of selective sweeps leave different signatures of local adaptation and our power to detect them will differ depending on which methods we use (Lange & Pool 2016). Some common approaches to scanning the genome for population-specific selective sweeps use F_{ST} (or F_{ST}-based) statistics to quantify genetic differentiation between populations. Local adaptation is expected to create genomic regions with more extreme differentiation than what would be expected under neutrality, since allele frequencies in these regions will change faster as the beneficial allele increases in frequency (Lewontin and Krakauer 1973). Neutral expectations can be inferred either with demographic simulations or an outlier approach. Demographic simulations, based on a previously estimated model of population history, can be used to mimic the history of the

populations being studied in the absence of natural selection. Outlier approaches rely on the genome-wide distribution of F_{ST} as a proxy for the neutral distribution, since neutral forces (including those due to demographic history) can broadly be expected to affect the whole genome similarly. Genome scans for regions under selection have typically focused on measuring F_{ST} or other statistics in windows of the genome of some predefined size to search for highly differentiated genomic regions.

A motivating empirical example for the present study comes from an investigation of the genetic basis of locally adaptive melanism in high altitude *Drosophila melanogaster* populations. Here, the authors used QTL mapping to identify genomic regions associated with derived dark pigmentation traits, and then used F_{ST} to scan these regions for signatures of selection (Bastide *et al.* 2016). One very narrow and strong QTL for highland Ethiopian melanism contained the well-known pigmentation gene *ebony*, which also contributed to melanic evolution in a Uganda population (Pool & Aquadro 2007, Rebeiz *et al.* 2009). Assessing genetic differentiation between the Ethiopia and Zambia populations for the window containing *ebony*, although full-window F_{ST} was only marginally elevated, it had a SNP with extremely high F_{ST} (0.85). Compared to demographic simulations, this window's maximum SNP F_{ST} value was among the top 1% of all windows, while its full-window F_{ST} was only among the 7% highest (Bastide *et al.* 2016). Simulated scenarios of soft sweeps from

standing variation replicated this pattern of extremely high maximum SNP F_{ST} and only moderately high window F_{ST}, suggesting that some kinds of selective sweeps that may not be detected using full-window F_{ST} could potentially be detected with a SNP-level F_{ST} approach. Further potential support for the use of SNP-level F_{ST} signals to detect adaptive events in this same species was demonstrated by much stronger parallel signatures of selection seen at the SNP level compared to the window level in fly populations that independently adapted to cold environments (Pool *et al.* 2017).

Challenges of using SNP-level F_{ST} values to detect selection include their variability due to random sampling effects (Weir *et al.* 2005) and the large number of tests that need to be made against a null distribution. Therefore, larger sample sizes are needed than for window F_{ST}. By using the highest SNP F_{ST} value within a window as a summary statistic for that window, and comparing it against null simulations with demography and recombination, we may somewhat improve the multiple testing issue, since here we are not treating all tightly linked SNPs as fully independent tests. Another advantage of this approach is that the maximum value summarizes each window of the genome, making it more comparable to any other window-based metric in terms of the number of tests and units of the genome analyzed. If full-window F_{ST} and maximum SNP F_{ST} are able to detect different types of selective events, then using both metrics could

result in a more comprehensive scan for signatures of local adaptation. The genome-wide distribution of these statistics in natural populations, compared to their neutral expectations, might also shed light on the contribution of different kinds of selective sweeps to local adaptation.

To understand the utility of using the highest F_{ST} value of any SNP within a window (hereafter F_{ST_MaxSNP}) as a local adaptation summary statistic, we performed power analyses based on extensive simulations, and then applied these results to empirical data from natural populations of *D. melanogaster*. We focused on comparisons between two populations and calculated the power of F_{ST_MaxSNP} to detect signatures of local adaptation under a wide range of different selective scenarios (including partial and/or soft sweeps) and demographic histories (including population bottlenecks and scenarios with ongoing migration). We performed demographic simulations and compared the power of F_{ST_MaxSNP} to both full-window F_{ST} based on all variable sites (herein, $F_{ST\_FullWin}$) and a comparative haplotype-based statistic (χ_{MD}, Lange & Pool 2016). Then, we investigated the genome-wide distribution of F_{ST_MaxSNP} and $F_{ST\_FullWin}$ among several natural populations of *D. melanogaster*, to determine whether either statistic was enriched genome-wide in empirical data compared to neutral expectations. Finally, we used an outlier approach to perform a genome scan for regions potentially under local adaptation between the Ethiopia and Zambia populations mentioned above, using F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD} (Lange & Pool 2016), and we

determined the extent of overlap between candidate regions identified according to these different methods. These analyses allowed us to both identify the parameter space in which F_{ST_MaxSNP} outperforms other statistics, and to assess the utility and complementarity of applying these approaches to real data.

Materials and Methods

Simulation power analysis

To generate adaptive and neutral distributions of genetic diversity, we performed simulations of demographic history scenarios with and without natural selection using *msms* (Ewing & Hermisson 2010). Our simulations consisted of two populations with a population split, and population-specific selective sweeps in the scenarios with natural selection. For each model, we obtained 10,000 replicates from which we calculated the statistics of interest. Power was calculated as the proportion of replicates under selection with a statistical value larger than 95% of the values obtained in its corresponding replicates without selection. We investigated the power of three different statistics: F_{ST_MaxSNP}, $F_{ST\_FullWin}$ and χ_{MD} (Lange & Pool 2016), which were calculated on windows of fixed size. F_{ST_MaxSNP} is based on the SNP within a window with the highest F_{ST} value. $F_{ST\_FullWin}$ was calculated as the ratio of the average between population variance for of all SNPs in a window over the average total (between +

within population) variance for all SNPs (Reynolds *et al*. 1983). No minor allele frequency filter was applied for SNP calling in the power analysis – but see below for criteria used to reject or accept any simulation replicate based on the allele frequency of the beneficial allele in particular. χ_{MD} stands for Comparative Haplotype Identity; it compares the average length of identical haplotypes in a window between two populations, and was calculated following Lange and Pool (2016). Our simulations used two general sets of parameters, following Lange and Pool (2016). One set with high effective population size (N_e = 2,500,000) was based on parameters from *Drosophila melanogaster* (with a population mutation rate of 0.01 and a population recombination rate of 0.05). The other set with a low N_e was based on parameters from humans (with population mutation and recombination rates of 0.001). To maintain similar scales of diversity and linkage between these scenarios, the default window size used in our simulations was 5,000 bp for simulations of populations with high N_e and 100,000 bp for simulations of populations with low N_e. The different window sizes for each population size reflect the amount of genetic diversity in high and low N_e populations. Except where otherwise stated, the sample size was 50 chromosomes.

We initially used scenarios of constant population size and a simple population split to simulate scenarios of selective sweeps with varying initial and final allele frequencies, representing hard and soft sweeps as

well as complete and partial sweeps. We also simulated scenarios of population bottlenecks and population splits for complete selective sweeps, and for scenarios with varying migration rates for hard sweeps (not constrained by ending allele frequency). For bottlenecks, the population that will experience local adaptation underwent a period of reduced population size for the first 0.01 coalescent units after the population split (which in most scenarios including these, occurred 0.05 coalescent units ago; Table S1).

The simulations of populations with high N_e were done for two different selection coefficients (s = 0.01 and s = 0.001) and simulations of populations with low N_e only included s = 0.01 (Table S1). Simulations of complete sweeps only used replicates in which the beneficial allele went to fixation. Simulations of partial sweeps only accepted replicates in which the beneficial allele stayed within 4% of the targeted ending frequency. Selection initiation time was adjusted in each case to maximize the proportion of accepted replicates. Moreover, in the scenarios with initial allele frequencies larger than $1/2N_e$, both the selected and non-selected populations had the same initial frequency.

For models that included migration (gene flow), selection of equal magnitudes but in opposite directions was imposed on each population. Per generation migration rates varied from 0.0004 to 0.004 in simulations with high N_e populations and from 0.01 to 0.10 in simulations with low N_e

populations. For each migration rate, split times varied from 0.1 to 1 coalescent unit.

We calculated the effect of sample size on the power of each statistic in six different scenarios: four models with demographic history of a simple isolation between two populations and two models with population size bottleneck. Of the simple isolation models, two models for high N_e populations were considered: one in which $F_{ST\_FullWin}$ outperformed F_{ST_MaxSNP} (initial allele frequency of $1/2N_e$ and final allele frequency of 0.4) and another where F_{ST_MaxSNP} outperformed $F_{ST\_FullWin}$ (initial frequency of 0.005 and final frequency of 0.7). Two scenarios for low N_e populations were also considered: one in which $F_{ST\_FullWin}$ outperformed F_{ST_MaxSNP} (initial allele frequency of $1/2N_e$ and final allele frequency of 0.5) and another where F_{ST_MaxSNP} outperformed $F_{ST\_FullWin}$ (initial frequency of 0.05 and final frequency of 0.8). For the bottleneck models, we used models with a bottleneck of 5% (*i.e.* a reduction to 5% of the prior N_e for 0.01 coalescent units in the adapting population immediately following the population split) and only models in which F_{ST_MaxSNP} outperformed the window wide statistics were considered: one model for high N_e population (initial allele frequency from 0.5% to 100%) and one for low N_e populations (initial allele frequency from 1% to 100%). For all the six scenarios, we used sample sizes of 10, 20, 50 (original sample size), 100, and 200 chromosomes.

We calculated the effect of window sizes on the power of each statistic in four different scenarios, the same scenarios of simple isolation used to calculate the power of sample sizes above. For the high N_e scenarios, we used window sizes of 5 kb (original size), 2 kb, 1 kb, 0.5 kb, 0.2 kb, 0.1 kb, and 1 bp. For the low N_e scenarios, we used window sizes of 100 kb (original size), 50 kb, 20 kb, 10 kb, 5 kb, 1 kb, and 1 bp. For the 1 bp (one single SNP) windows, we only calculated F_{ST} (here $F_{ST\_MaxSNP} = F_{ST\_FullWin}$). To calculate χ_{MD}, we used a minimum haplotype threshold of 10% of the window size (as was used for the original analyses). For each window size smaller than the original, we applied a p-value Bonferroni multiple testing correction proportional to the reduction in size (or equivalently, the increased number of windows needed to cover a given genomic region) to calculate power. That is, while for the standard window size power is the number of replicates with a p-value of 0.05 or lower, for a window half the size of the original the p-value would need to be 0.025 or lower. Except for the window size of 1 bp, in which the correction was the average number of SNPs in the window with the largest size (the default window size used in our other analyses).

Empirical enrichment of F_{ST_MaxSNP} *and* $F_{ST\_FullWin}$ *- data and simulations*

Our data set consists of individual fly strain genomes from six natural populations of *D. melanogaster*: one non-human commensal population

from Kafue, Zambia (KF) and five human commensal populations from different countries: Zambia (ZI), South Africa (SD), Rwanda (RG), Ethiopia (EF) and France (FR), using data from Lack *et al.* (2016) and Sprengelmeyer *et al.* (2020). From each population, for each chromosome arm (ChrX, Chr2L, Chr2R, Chr3L, Chr3R), we excluded genomes from lines with a known inversion for that arm. To boost the sample size of two populations with genomes from partially inbred lines (Ethiopia and France), instead of only using homozygous regions of the genome (as in the original filtering of the published data set) we also included heterozygous regions identified by Lack *et al*. (2016), and therefore counted two alleles at each site from these regions. For any pair of lines with excess identity by descent (IBD) between them (defined as more than 10 megabases of IBD outside previously defined regions of low recombination; Lack *et al.* 2016), we excluded one member of the pair from this data set. Non-African admixture was filtered out from haploid data from African populations based on data from Lack *et al*. (2016). For each population sample and each chromosome arm, we chose a sample size to jointly maximize the number of analyzable sites and the sample size itself. Our resulting sample sizes are shown on Table S2. For sites with more than that number of alleles called, we downsampled to match the chosen sample size.

We calculated pairwise $F_{ST\_FullWin}$ and F_{ST_MaxSNP} for all populations using diversity-scaled window sizes designed to contain 250 non-singleton SNPs

in the ZI sample. F_{ST_MaxSNP} and $F_{ST\_FullWin}$ were calculated using each SNP with minor allele count larger than two, using the same approach described in the power analysis. To compare empirical and null distributions for similar recombination rates, each window was assigned to one of five recombination rates bins based on estimates from Comeron *et al*. (2012); the bins corresponded to recombination rates from 0.5-1, 1-1.5, 1.5-2, 2-3, and greater than 3. Windows with recombination rates lower than 0.5 were not used due to low spatial resolution for localizing signatures of selection in low recombination regions. We obtained p-values for each window using neutral demographic simulations performed using *ms* (Hudson 2002). Demographic simulations were performed using parameters estimated for the evolutionary history of nine populations of *D. melanogaster*, including all the populations we analyzed (Sprengelmeyer *et al.* 2020). The other three populations were lowland Ethiopia (EA), Cameroon (CO), and Egypt (EG). We did not use those three populations in our empirical analyses due to their lower sample sizes. Nonetheless, they were included in the simulations in order to accurately reflect the estimated patterns of migration.

Each demographic model had been estimated based on tentatively neutral genetic markers (short introns and 4-fold synonymous sites from regions with sex-averaged recombination rates of at least 1 cM/Mb) from inversion-free chromosome arms (Sprengelmeyer *et al.* 2020). A model

was estimated for each of three chromosome arms that had lower inversion frequencies (X, 2R, and 3L), and the history was inferred iteratively, such that not all population samples were present in the same model. To better approximate genetic diversity in all populations, we used two sets of demographic models: Northern model (containing ZI, RG, CO, EF, FR, EG, EA) and Southern model (containing ZI, RG, CO, SD, and KF). The Northern model for the chromosome X was subdivided into two sub-models (one with ZI, RG, CO, EF, EA and another with ZI, RG, CO, FR, EG). Hence, we simulated four Northern models and three Southern models (command lines in Table S2). The models for the autosomal chromosome arms (2R and 3L) were simulated using the highest sample sizes for any autosomal arm of each population (Table S2). Simulated sample sizes were downsampled to match the sample sizes of each specific arm when comparing empirical and simulated F_{ST} patterns for any given arm. A minor allele count of three or greater was also applied to the simulated data, mimicking the same filtering used on the empirical data. The window size and crossing over rate used in each replicate were based on a random sampling with replacement from the empirical windows, and the single gene conversion rate and mean tract length were based on the estimates of Comeron *et al*. (2012). Therefore, a null distribution was generated for each model and each recombination bin (described above).

For each model and each recombination bin, 50,000 replicates were simulated.

Enrichment calculation

$F_{ST\_FullWin}$ and F_{ST_MaxSNP} were calculated for each population pair and each chromosome arm. F_{ST} was calculated for the simulated data using the same sample sizes as the empirical data (Table S2). For sites with more than two alleles, only the two most common alleles were kept. Sites with minor allele counts lower than two were discarded from empirical and simulated analyses.

P-values were calculated for each window based on the neutral distribution of its corresponding recombination group. Windows from chromosome X were compared to neutral distributions based on the model for chromosome X. For autosomal loci, we determined that simulations from the 3L model yielded somewhat milder outlier enrichments than the 2R model, and therefore we conservatively focused on results from the 3L model.

We calculated p-value enrichments for $F_{ST\_FullWin}$ and F_{ST_MaxSNP} using p-value bins of width equal to 0.05, resulting in 20 bins of p-value 0 to 1. We counted how many windows had a given p-value for each bin and divided the observed number by how many windows we expected to have with a p-value in that bin based on simulated data. Neighboring windows with low

p-value could be showing the effect of a single selective sweep. Therefore, we complemented this outlier window enrichment analysis with one based on "outlier regions". We intentionally defined outlier regions generously, preferring to falsely lump two sweeps versus splitting a single sweep into two or more regions. Formally, starting with the window containing the lowest p-values, we extended the region surrounding it until we reached a stretch of five consecutive windows with $p > 0.1$ to create an outlier region. We removed the outlier regions from our analysis and repeated the process until the signal of enrichment was erased (defined as the $p < 0.05$ bin having no more enrichment than the $0.05 < p < 0.1$ bin). For each of F_{ST_MaxSNP} and $F_{ST\_FullWin}$, we recorded the total number of outlier regions that had to be removed for a given population pair. On the other hand, since neighboring windows with high p-values (low F_{ST}) could be showing shared sweeps, we repeated the process described above for outlier regions based on high p-values. For high p-value windows, we defined enrichment as the $p > 0.95$ bin having no more windows than the $0.9 < p < 0.95$ bin.

Genome scan for regions under selection - Ethiopia vs. Zambia

We performed a genome scan for candidate regions under selection between the Ethiopia (EF) and Zambia (ZI) populations. We calculated $F_{ST\_FullWin}$, F_{ST_MaxSNP}, and χ_{MD} for each window of the genome. We used an outlier approach and considered windows in the top 1% of each statistic to be the

candidate regions under selection. Here, we combined multiple outlier windows into the same outlier region if they were separated by no more than five windows with p-value > 0.01. To investigate whether the candidate regions detected with each statistic were the same or unique, we calculated how many regions overlapped between the different statistics. We considered that two regions were overlapping if at least 50% of the smaller region overlapped the larger one.

For each list of candidate regions under selection, we performed a GO term enrichment analysis using a method initially described by Pool *et al.* 2012. For each gene within a candidate region, we obtained GO term annotations from FlyBase. The GO terms for each gene also included all the parents of each term. GO terms that appeared repeatedly in a candidate region were counted only once for that region. We calculated the p-values for each GO term based on 10,000 permutations of the genomic locations of the outlier regions. This procedure allows genes to have different null probabilities of being outliers, particularly based on their length. We obtained a list of enriched GO terms for each statistic defined as the GO terms with raw p-values less than or equal or to 0.01. We then determined the overlap between the three lists of enriched GO terms.

To investigate whether F_{ST_MaxSNP} and $F_{ST\_FullWin}$ outliers might be detecting different kinds of selective sweeps, we focused on the outlier regions that

were exclusive to each statistic. We calculated the frequency of the most common haplotype, haplotype homozygosity, and the H2/H1 statistic (Garud *et al*. 2015) for the window with the most extreme statistic in each region. In case of ties, one window was chosen randomly (for F_{ST_MaxSNP}, randomizations were proportional to the number of top SNPs in each window). The expectation is for hard sweeps, a single haplotype has risen in frequency in the population, and therefore the frequency of the most common haplotype, as well as haplotype homozygosity, should be higher following a hard sweep than a soft sweep. H2/H1 (calculated following Garud *et al*. 2015) calculates the ratio of the haplotype homozygosity calculated without the most common haplotype (H2) over the overall haplotype homozygosity including the most common haplotype (H1); it should be higher following soft sweeps than hard sweeps. We calculated these statistics for all windows of the genome with recombination rates above 0.5 that had a minimum sample size of 10 chromosomes from each population. For each window we, excluded haplotypes with an amount of missing data above the average for that window. We did not consider sites with singletons (only one of the haplotypes had a different allele for that site) when calculating haplotype frequencies. Ambiguous haplotypes were assigned to a matching haplotype; the assignment probability for each matching haplotype was proportional to its frequency.

To investigate whether the sites with highest F_{ST} values in the outlier genomic regions for F_{ST_MaxSNP} potentially were the targets of selection, we calculated their enrichment across different categories of functional sites. We classified each site into five classes: nonsynonymous, synonymous (only considering fourfold synonymous), untranslated regions of the mRNA (UTR), intronic, and intergenic. For each outlier region, we focused on the SNP(s) with the highest F_{ST} value. If more than one site were tied for highest F_{ST} in an outlier region, instead of counting 1 for each site class we counted 1/(the number of top sites), so the total count for each region was always 1 regardless of how many SNPs were tied for highest F_{ST} value. We then counted how many sites in each class were present across all outlier regions. We also calculated the genome-wide proportion of each site class, restricting our analysis to sites in which the average minor allele frequency between the Ethiopia and Zambia populations were within the range of average minor allele frequency for all sites with the highest F_{ST} values in the outlier regions. Lastly, we calculated enrichment for each site class as the ratio between the proportion of sites in the outlier regions over the proportion of sites in the genome.

Results

Maximum SNP F_{ST} *and full-window summaries have complementary power to detect local adaptation*

We performed power analyses of F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD} using population genetic simulations with and without natural selection. We used *msms* (Ewing & Hermisson 2010) to simulate a two-population isolation model with positive selection in one population but not the other. with constrained initial and final allele frequencies, yielding local sweeps that could be hard or soft, and partial or complete. Beyond the simple isolation model, demographic scenarios with population size bottlenecks or migration were simulated as well (simulation commands in Table S1). For each scenario, we simulated both a low effective population size (N_e) model with mutation and recombination parameters based on estimates for humans, and a high N_e model with parameters motivated by *Drosophila melanogaster* (see Materials and Methods), following the design of a previous power analysis study that did not include F_{ST_MaxSNP} (Lange & Pool 2016). These low and high N_e scenarios entail very different levels of diversity and scales of linkage disequilibrium (motivating contrasting window sizes of 100 kb versus 5 kb in most of our analyses), and they may therefore provide useful reference points for a range of taxa beyond the motivating species themselves. For the low N_e simulations, we focused on sweeps with a selection coefficient of s = 0.01. In high N_e species, many

successful sweeps may have weaker advantages. Here, we focused on results with s = 0.001. High N_e results with s = 0.01 gave similar results except where noted below (Supplementary Table 1). F_{ST_MaxSNP}, $F_{ST\_Fullwin}$, and χ_{MD} were calculated between the selected and non-selected populations at the end of the simulation. Power was defined in a locus-specific context, based on the proportion of selection simulations giving a more extreme value of the summary statistic than the 95th quantile of its distribution from neutral simulations.

Unsurprisingly, all three statistics were found to have high power for the case of complete hard sweeps (Figure 1; Table S1). These simulations were conditioned on fixation of a beneficial new mutation in one population that had not occurred in the other population. In light of this fixed difference, F_{ST_MaxSNP} in all replicates had its maximum value (F_{ST_MaxSNP} = 1). In such cases, the power of F_{ST_MaxSNP} was binary, either zero or one, depending on whether or not 5% of the corresponding neutral replicates had an allele that reached fixation. In our simple isolation model, the likelihood that a neutral allele can reach fixation increases with the split time (Table S1; Figure S1). Stronger bottlenecks also boost the likelihood of having neutral alleles reach fixation (Table S1; Figure S2, Figure S3). Hence, power for F_{ST_MaxSNP} to detect complete hard sweeps goes from high, for recent splits and weaker bottlenecks, to zero for histories in which more than 5% of neutral replicates contain a fixed difference. Similarly, $F_{ST\_FullWin}$ and χ_{MD} had higher

power to detect signatures of local adaptation following recent splits and in weaker bottlenecks, but their change in power was gradual and continuous instead of binary.

In the case of complete or nearly complete soft sweeps, F_{ST_MaxSNP} showed a clear power advantage over $F_{ST\_FullWin}$ and χ_{MD}. Notably, for sweeps ending between 80% and 100% frequency, F_{ST_MaxSNP} had high power to detect local adaptation, even for cases with rather high initial frequencies of the beneficial allele (e.g. 10%; Figure 1; Figure 2). In contrast, $F_{ST\_FullWin}$ and χ_{MD} showed rapidly diminishing performance as sweeps became softer (Figure 1; Figure 2). These results make sense, in that beneficial alleles that drift to higher pre-selection frequencies have more time to recombine onto multiple haplotypes, and recombination events will have happened closer to the selected site on average. Therefore, soft sweeps are generally narrower in width and may not substantially alter full-window statistics (Catania *et al.* 2004, Schlenke & Begun 2004, Hermisson & Pennings 2005). Although the two full-window statistics maintained good power for lower initial frequencies, some of the replicates of those scenarios are actually generating hard sweeps due to the chance survival of a single haplotype carrying the favored variant (Jensen 2014), as shown by an average number of beneficial haplotypes lower than two in these simulations (Figure 2). Moreover, as the average number of haplotypes

carrying the favored variant increased, the power of the full-window statistics decreased (Figure 2), while the power of F_{ST_MaxSNP} was unchanged.

Contrasting results were obtained for partial, harder sweep scenarios. In cases where new mutations or rare standing variants were only boosted to intermediate frequencies, $F_{ST\_FullWin}$ and χ_{MD} had fairly strong power, whereas F_{ST_MaxSNP} declined sharply in effectiveness at around 60% final frequency for hard sweeps (Figure 1). These results are also intuitive, in that partial hard sweeps can meaningfully alter allele frequencies across a whole window and generate a class of identical haplotypes, even though no single SNP traverses an extreme range of frequencies. The broadly similar power profiles of $F_{ST\_FullWin}$ and χ_{MD} are somewhat surprising in light of their distinct basis (albeit consistent with Lange and Pool, 2016). Less surprising is that for the challenging scenario of partial soft sweeps, none of the three statistics showed strong power in the scenarios examined (Figure 1).

Whereas the above simulations had no migration, we also wondered if F_{ST_MaxSNP} might prove useful in detecting targets of local adaptation for which genetic differentiation had been whittled down in width by recombination with migrant alleles over time (Sakamoto & Innan 2019). We therefore simulated scenarios with varying combinations of migration rate and population split time, while assuming symmetric migration rates and equal but opposing selective pressures. Overall, F_{ST_MaxSNP} and $F_{ST\_FullWin}$ performed very similarly to each other and better than χ_{MD}. Particularly in

the high N_e scenarios (which feature a higher ratio of recombination to mutation events) with intermediate migration rates, there was a narrow space of parameters in which F_{ST_MaxSNP} performed slightly better than $F_{ST\_FullWin}$ (Figure S4). The split time between the populations greatly affected the power of χ_{MD}, which performed better on recent splits. The power of the F_{ST} statistics showed a small improvement for more recent splits and intermediate migration rates. Although small, the effect of split time also seemed more pronounced on $F_{ST\_FullWin}$ than F_{ST_MaxSNP} (Figure S4). Overall, these analyses provide only modest support for the notion that F_{ST_MaxSNP} could help detect peaks of genetic differentiation driven by local adaptation that have been narrowed by migration and recombination.

In the above simulations, we used a sample size of 50 chromosomes per population. We generally expect statistical power to be correlated with sample size and understanding the effect of sample size on the power of each statistic is relevant when designing an experiment or choosing which statistics to use. We analyzed the power of F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD} in three scenarios for high N_e and three for low N_e. We chose scenarios in which F_{ST_MaxSNP} and the window wide statistics performed differently: a mostly complete soft sweep, a complete soft sweep with a bottleneck, and a partial hard sweep. We found that sample size had a stronger effect on F_{ST_MaxSNP} than on the window wide statistics (Figure 3). F_{ST_MaxSNP} is based on allele frequencies at a single site, so it is more sensitive to the increased sampling

variance at lower sample sizes than window wide statistics. The sampling variance in each SNP in a window should fluctuate around the mean, so when information from each SNP is combined the full-window F_{ST} statistic suffers less from the reduced sample size. Demographic history also affected the effect of sample size on each statistic: in scenarios with a population bottleneck, which also increases sampling variance, the power of F_{ST_MaxSNP} changed from near 1 at sample size 50 or higher to 0 at sample sizes smaller than 50 (Figure 3C, 3D). More generally, F_{ST_MaxSNP} was found to perform much better with 50 chromosomes than with 20, but showed relatively less improvement for sample sizes larger than 50.

We also analyzed the effect of window size on the power of each statistic, with the aim of determining whether there would be a window size for which a single statistic would perform well in contrasting scenarios. For example, one might hope that $F_{ST\_FullWin}$ for a narrower window might retain good performance for partial hard sweeps, while also capturing the advantages of F_{ST_MaxSNP} for complete soft sweeps. We explored four scenarios of partial sweeps, two for the high N_e and two for the low N_e. For each population size, we chose one scenario in which the power of F_{ST_MaxSNP} outperformed $F_{ST\_FullWin}$ and χ_{MD}, and one in which it underperformed. In practice, a reduction in window size would result in an increase in the number of tests performed in a genome scan. Therefore, we applied a Bonferroni correction to the p-value proportional to the reduction in size.

The correction for window size equal to one site (a single SNP) was proportional to average number of SNPs in the largest window (the default window size used in our analyses). Our results showed that, for the two scenarios in which F_{ST_MaxSNP} outperformed $F_{ST\_FullWin}$ and χ_{MD}, the power of each statistic remained mostly constant (Figure 4). For the scenarios in which $F_{ST\_FullWin}$ and χ_{MD} had an advantage, the power of each statistic, as well as the difference among them, declined with smaller window sizes. Overall, there was no window size in which a single statistic performed well for all scenarios, and hence it may be preferable to apply F_{ST_MaxSNP} and full-window statistics separately to empirical data.

Outliers for F_{ST_MaxSNP} *and* $F_{ST\_FullWin}$ *are enriched in empirical data*

In light of the above results, we were interested in applying both F_{ST_MaxSNP} and $F_{ST\_FullWin}$ to an empirical data set, in part with an interest in quantifying the relative enrichment of outliers for each statistic and what that might hint about the modes of selection active in these populations. We chose to focus on data from the *Drosophila* Genome Nexus (Lack *et al.* 2015, 2016), because it contained several populations of *D. melanogaster* that were linked by an estimated model of population history (Sprengelmeyer *et al.* 2020) and had at least minimal sample sizes available for studying genome-wide patterns of F_{ST} (Table S2). We included six natural populations of flies. From the ancestral range in Zambia, we included one

town population (Siavonga) and one wilderness population (Kafue). We also included four additional town populations: from Rwanda, South Africa, Ethiopia, and France (the latter three having independently colonized colder environments; Pool *et al*. 2017).

We calculated a p-value for each empirical window in each pairwise population comparison, based on neutral distributions of F_{ST_MaxSNP} or $F_{ST\_FullWin}$ generated using coalescent simulations of the estimated demographic history (Sprengelmeyer *et al.* 2020; simulation commands in Table S2). Under neutrality, a uniform distribution of p-values is expected. In general, for most population pairs, the distribution of p-values for F_{ST_MaxSNP} and $F_{ST\_FullWin}$ showed a U-shape instead of a uniform distribution (e.g. Figure 5A-B). The deviation from the expected uniform distribution could be attributed to the action of natural selection producing windows with higher and lower F_{ST} than expected (*e.g.* by local adaptation and shared sweeps respectively) or by a misspecification of the neutral demographic model. However, average F_{ST} values of simulated data from this model were found to align well with empirical measurements (Sprengelmeyer *et al*. 2020), and similar results were found with other summary statistics. The enrichment of high F_{ST} (defined as p-values from 0 to 0.05) and low F_{ST} (p-values from 0.95 to 1) varied for each statistic and across the population comparisons (Figure 5C-D). Particularly for high $F_{ST\_FullWin}$, the strongest enrichments were often observed for more geographically proximate, closely related

population pairs, perhaps reflecting reduced noise from neutral genetic differentiation.

All population pair comparisons showed an enrichment for windows with high $F_{ST\_FullWin}$. The smallest enrichment was found between the Zambia (town) and France populations, for which there were 3.29 more windows with high $F_{ST\_FullWin}$ than expected by chance. The highest enrichment was found in the comparison between the South Africa and Kafue (Zambia wilderness) populations, with an enrichment factor of 9.06. For F_{ST_MaxSNP}, eight population pairs had an enrichment value > 2, the highest being 5.41 (between the Zambian town and wilderness populations, and between South Africa and Rwanda). On the other hand, one population pair was slightly depleted of windows with high F_{ST_MaxSNP} (enrichment to 0.87 between France and Ethiopia). In nearly all comparisons, $F_{ST\_FullWin}$ showed higher enrichment than F_{ST_MaxSNP} (Figure 5). However, this difference in enrichment could be influenced by single local sweeps that generate multiple linked outlier windows. We therefore pursued a complementary analysis in which nearby outlier windows were merged into "outlier regions", which were then removed one at a time until the observed enrichment was erased (see Materials and Methods). For almost every population pair, we had to remove a larger number of regions to erase the signal of enrichment of $F_{ST\_FullWin}$ than the signal of F_{ST_MaxSNP} (Figure 5E-F). Hence, the greater enrichment of $F_{ST\_FullWin}$ relative to F_{ST_MaxSNP} does not appear to be a product of

broader linkage signals of $F_{ST\_FullWin}$ outliers alone. Instead, this pattern could hint that sweeps in the unique detection parameter space of $F_{ST\_FullWin}$ (*i.e.* partial harder sweeps) are more common among these populations than sweeps in the unique space of F_{ST_MaxSNP} (*i.e.* more complete softer sweeps). However, these results may be influenced by other evolutionary forces as well, and they do not offer definitive conclusions about the prevalence of different models of selection (see Discussion).

Our simulation results above suggested that high F_{ST_MaxSNP} and $F_{ST\_FullWin}$ outliers might be capturing different kinds of selective sweeps. To assess this possibility from the empirical data, we focused on high F_{ST_MaxSNP} and $F_{ST\_FullWin}$ outlier regions (as described above) from the Ethiopia *vs.* Zambia comparison. We calculated the frequency of the most common haplotype, haplotype homozygosity, and the H2/H1 statistic (Garud *et al*. 2015) for the outlier regions exclusively detected with F_{ST_MaxSNP} and those exclusively detected with $F_{ST\_FullWin}$. Congruent with F_{ST_MaxSNP} exclusive outliers mainly detecting cases of soft sweeps and $F_{ST\_FullWin}$ exclusive outliers detecting hard partial sweeps, we found that for both the Ethiopian and the Zambian populations, the frequency of the most common haplotype and haplotype homozygosity was lower in the F_{ST_MaxSNP} outliers, while H2/H1 was higher (meaning the haplotype homozygosity calculated with and without the most common haplotype was more similar to each other) in the F_{ST_MaxSNP} exclusive outliers than $F_{ST\_FullWin}$ (Figure S5).

We also performed an outlier removal analysis for windows with high p-values (low F_{ST}), which could reflect shared sweeps or other processes. Similar to the low p-value enrichment analysis, we found varied results for each statistic and population pair (Figure S6).

Genome scan for signatures of selection

We chose to complement the above multi-population analysis of genome-wide patterns with a closer analysis of a single population pair. We chose to compare the Ethiopia and Zambia town populations because (1) Their relatively large sample sizes of 129-181 and 60-76 respectively for each chromosome arm (Table S2) are more conducive to the analysis of specific F_{ST_MaxSNP} outliers, (2) These populations showed enrichments of both F_{ST_MaxSNP} and $F_{ST\_FullWin}$ (Figure 5), and (3) Past results from these populations helped motivate the present study (*e.g.* Bastide *et al.* 2016). We performed genome scans for regions potentially under population-specific selection between these populations using F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD}. For each statistic, we obtained a list of outlier windows (top 1%), and as above, we merged nearby outlier windows into regions (Materials and Methods). We obtained 138 outlier regions for F_{ST_MaxSNP}, 138 for $F_{ST\_FullWin}$, and 155 for χ_{MD}. Our results showed an overlap of just 39% between the outlier regions detected with F_{ST_MaxSNP} and $F_{ST\_FullWin}$. Perhaps surprisingly in light of the above power results, there was a smaller overlap of either F_{ST} metric with χ_{MD} (Figure 6A),

although the overlap of the haplotype statistic with $F_{ST\_FullWin}$ was indeed slightly greater. In regions that were outliers for F_{ST_MaxSNP} but not $F_{ST\_FullWin}$, the distribution of individual SNP F_{ST} values often had a narrow sharp F_{ST} peak, with most of the other SNPs having low F_{ST} values. On the contrary, in regions there were outliers for $F_{ST\_FullWin}$ but not F_{ST_MaxSNP}, often no single SNP had a large F_{ST} value, but there was a broad moderate F_{ST} plateau with many SNPs showing intermediate F_{ST} values (Figure 7).

The SNP with the highest F_{ST} value in each outlier region for F_{ST_MaxSNP} could potentially represent the target of selection; therefore we asked whether they were enriched for functional site annotations generally associated with greater evidence for positive and negative selection. We classified these SNPs into five different classes: nonsynonymous, synonymous, untranslated region of the mRNA (UTR), intronic, and intergenic. We then compared the proportion of "top SNPs" (*i.e.* having the highest SNP F_{ST} within a F_{ST_MaxSNP} outlier region) in each functional site category against that category's genome-wide proportion, based on SNPs with similar allele frequencies. We found the biggest enrichment among nonsynonymous (protein-altering) sites, with an enrichment of 3.2, followed by UTR sites (Figure 8). The remaining classes were not enriched, and the intronic class was the most depleted class, with an enrichment of 0.8 (Figure 8). Previous studies have found evidence of selection on noncoding sites in *Drosophila*, especially on UTR sites - which have shown

more selective constraints and proportionally more adaptive substitutions than intronic and intergenic sites (Andolfatto 2005, Lange & Pool 2018). The enrichment of nonsynonymous and UTR sites in our analysis also mirrors results from human F_{ST} outliers (Barreiro *et al.* 2008). Overall, there is a strong tendency for our top SNPs to occur in site categories more likely to affect fitness, as we would predict if some of them are actual targets of selection. If a beneficial mutation in these sites was already present as standing variation in the population before the onset of selection, the increase in frequency of beneficial mutation in a single population could result in a narrow sharp F_{ST} peak within the genomic region (Figure 7).

We then performed Gene Ontology (GO) term enrichment analysis separately for each statistic's list of outlier regions. Considering only GO terms with raw p-value < 0.01 from each list, we found mostly lower overlaps between enriched GO terms compared to the spatial overlap between outlier regions (Figure 6B; Table S3). The three statistics differed substantially in the number of enriched GO terms by this criterion: 357 for $F_{ST\_FullWin}$, 133 for F_{ST_MaxSNP}, and 71 for χ_{MD} (out of 47,496 total GO terms tested). We emphasize that enriched terms in each set are not necessarily independent and any given list of enriched GO terms will contain overlapping categories. The relative overlap between GO terms enriched for each statistic largely followed the relative numbers of enriched GO

terms for each (Figure 6B). Mirroring the outlier region results, most enriched GO terms were detected for only one of the three statistics, highlighting the complementarity of each statistic described above. Different categories of genes have different mutational target sizes and may also vary in their ability to harbor potentially functional variation. Hence, the supply of standing genetic variation to generate soft (as opposed to hard) sweeps may differ between GO categories, as hinted by our results. Here, a number of the most enriched GO terms for $F_{ST\_FullWin}$ involved nucleotide/ribonucleotide binding (Table S3). Whereas, many of the most enriched GO terms for F_{ST_MaxSNP} pertained to ion channels, a finding concordant with previously-reported parallel signals of positive selection in cold-adapted *D. melanogaster* populations, based on SNP-level genetic differentiation outliers (Pool *et al.* 2017).

Discussion

F_{ST_MaxSNP} *complements other statistics by detecting soft sweeps*

Identifying regions under selection can help us answer further questions about the evolution of local adaptation, such as which biological functions are under selective pressure, the number of loci underlying adaptive events, the source of the adaptive variation, and the kinds of genetic changes that might be under selection. Our results underscore the importance of deploying methods capable of capturing different kinds of

selective sweeps when the aim of the study is to identify as many genes potentially under local adaptation as possible.

F_{ST_MaxSNP} in particular, seems to be especially useful to detect soft sweeps with relatively large initial and final frequencies of the beneficial allele. Instances of mostly complete soft sweeps, as simulated here, represent regions in which a beneficial allele was present in several different haplotypes that might have increased in frequency along with the beneficial allele. While the selected SNP itself changed in frequency drastically, resulting in a large F_{ST_MaxSNP}, the alleles around it must have changed in frequency to a lesser degree because many background haplotypes were hitchhiking along with the beneficial allele. Therefore, while the beneficial variant can have an extreme F_{ST} value, the lower allele frequency changes in the other SNPs in that window would result in a $F_{ST\_FullWin}$ that is not statistically significant, and thus a low power to detect a selective sweep under these conditions.

The full-window metrics, $F_{ST\_FullWin}$ and χ_{MD}, had greater power than F_{ST_MaxSNP} to detect relatively harder, partial sweeps that had intermediate final allele frequencies. In these sweeps, no individual SNP changed dramatically in frequency, so none have F_{ST} values higher than what could be obtained randomly in the genome. However, the increase in frequency of one or a few haplotypes resulted in many SNPs in the same region with intermediate F_{ST}, producing a window-wide pattern that is too extreme to

be generated by chance - even if each single marker individually did not have an extreme F_{ST} value.

We note that Kimura *et al*. (2007) also compared the power of a maximum SNP F_{ST} statistic against a haplotype statistic, in the context of detecting hard sweeps from SNP genotyping data. Consistent with our study, they found that the haplotype statistic performed better than maximum SNP F_{ST} in this hard sweep context. They also found that among simulation replicates, these two statistics were inversely correlated. These results are congruent with our general findings of complementary power between maximum SNP F_{ST} and either a comparative haplotype identity statistic or a full-window F_{ST} statistic.

The power of each statistic depends on population history

Importantly, the relative utility of each statistic to detect local adaptation was found to vary as a function of demographic history. For example, although F_{ST_MaxSNP} is generally much better than the studied full-window statistics at detecting complete soft sweeps, this advantage can be reversed if demography, in conjunction with sample sizes, yields fixed differences in at least 5% of windows under neutrality (in which case the power of F_{ST_MaxSNP} as we have defined it becomes zero). We demonstrated this phenomenon in cases with elevated genetic drift between populations, resulting from either a more ancient population split (Figure S1) or else a

strong population bottleneck in the adapting population (Figure S2; Figure S3). These results underscore the importance of performing simulations to test whether F_{ST_MaxSNP} is expected to be a useful metric for any given population pair of interest.

There was little difference in the power of F_{ST_MaxSNP} and $F_{ST\_FullWin}$ to detect regions under selection in scenarios with varying migration rates. We had wondered if F_{ST_MaxSNP} would outperform $F_{ST\_FullWin}$ in scenarios with older splits, as selection might only maintain a narrow window of differentiation between the two populations in the presence of long-term recombination with migrant haplotypes (Sakamoto & Innan 2019). Nonetheless, differences in the time of population divergence and local adaptation only had a small effect in a very narrow space of parameters (intermediate migration rates for high N_e populations, Figure S1), suggesting that even in scenarios with recent divergence, the populations had already reached a state of equilibrium and the balance between migration, selection, and recombination, which did not result in contrasting signatures of selection between F_{ST_MaxSNP} and $F_{ST\_FullWin}$. However, both metrics outperformed χ_{MD} on the simulated scenarios, indicating that selection could not maintain long shared haplotypes in the presence of migration.

For simplicity, we have limited our focus to the detection of local adaptation from two-population isolation models (with and without migration). Such histories may be generally relevant for many taxa,

including species that have recently invaded novel ranges, comparisons between domestic organisms and wild relatives, and island-dwelling taxa. Still, it is worth keeping in mind that many species exist as geographically complex mosaics of populations connected by migration. Patterns of genetic variation produced by positive selection (and by neutral processes) in spatially explicit contexts involve additional nuance not reflected in our study (*e.g.* Ralph & Coop 2015, Lee & Coop 2017). For example, a hard sweep in a subdivided population is expected to be narrower than it would otherwise be, as recombination events continue to whittle down the sweeping haplotype as it spreads from one deme to another (Santiago & Caballero 2005), which might further support the analysis of F_{ST} at the level of SNPs or narrower windows. However, more detailed study is needed to fully document the expected genomic scale of F_{ST} outliers in spatially complex population models.

Consideration must be given to window size, sample size, and multiple testing

In this study, we have used neutral demographic simulations to estimate statistical power at the single window level, only penalizing multiple tests when comparing between window sizes. Clearly, our results do not imply the power to identify genome-wide significant loci, which is only rarely attainable for population genomic scans. Instead, most genome

scans aim to identify good candidates for downstream study, and our results are best interpreted in terms of the relative utility of these summary statistics to identify local adaptation candidates. Similar interpretations should apply to genome scan outliers based on F_{ST_MaxSNP} versus other window-based summary statistics, unless it can be shown (*e.g.* via neutral demographic simulations) that an extreme observed value of F_{ST_MaxSNP} would not be expected anywhere in the genome.

In light of the complementary performance of F_{ST_MaxSNP} and $F_{ST\_FullWin}$ for the non-migration cases, we tested whether $F_{ST\_FullWin}$ across shorter windows could yield a balance of reasonable power to detect both complete soft sweeps and partial hard sweeps. However, the relationship between window size and the power - while accounting for the increase in the number of tests in smaller windows - did not follow this prediction. Our results suggest that applying both F_{ST_MaxSNP} and $F_{ST\_FullWin}$ to conventionally-sized windows is preferable to shrinking the window size in an effort to identify narrower soft sweeps. Nevertheless, window size remains a challenging decision in genome scans including those searching for local adaptation. Importantly, the scale of elevated genetic differentiation depends on multiple factors, including the magnitudes of selection, recombination, and migration, the timing of the onset of adaptation, and as we highlight, the initial frequency of a favored variant. In general, we suggest that genetic differentiation on both SNP and broader scales should be incorporated

into scans for local adaptation, whether using the specific summary statistics described here, or attempting to develop a single statistic or integrated analysis framework that encompasses the advantages of both.

An important caveat of using F_{ST_MaxSNP} is that it requires a greater sample size than $F_{ST\_FullWin}$. With smaller samples, it is easy to get a large F_{ST_MaxSNP} at one of the many analyzed SNPs through sampling variance alone, whereas an extreme $F_{ST\_FullWin}$ value is less likely in this scenario. It is difficult to provide any universal advice regarding sample size, because the neutral variance of F_{ST_MaxSNP} also depends strongly on demographic history, as shown above. Nonetheless, we have shown that in two scenarios in which F_{ST_MaxSNP} outperformed $F_{ST\_FullWin}$, its power declined considerably when we decreased the sample size from 50 to 20 chromosomes. Although the relationship between sample size and power will depend on the specific populations being studied, the utility of F_{ST_MaxSNP} seems most promising when sample sizes are around 100 alleles per population or more. However, it would be advisable to conduct neutral simulations based on estimated or suspected demography, in order to identify sample sizes for which it is very unlikely to get extreme single-SNP F_{ST} values in the absence of local adaptation.

Both $F_{ST\_FullWin}$ *and* F_{ST_MaxSNP} *outliers are enriched among* Drosophila *populations*

When we applied $F_{ST\_FullWin}$ and F_{ST_MaxSNP} to empirical data from *D. melanogaster* populations, we found that enrichment patterns of $F_{ST\_FullWin}$ and F_{ST_MaxSNP} varied among population pairs, both for high and low F_{ST} values. The excess of windows with high F_{ST} observed could be explained by local adaptation: unique selective sweeps in one population increase the differentiation between two populations in that region. Not all population pairs showed the same degree of enrichment for high F_{ST}. A larger enrichment could be due to a higher number of selective sweeps between two populations, stronger selective events that impacted a larger region of the genome, or a neutral history more conducive to outlier detection. The populations we studied cover a large geographical scale, most are located in sub-Saharan Africa and one in Europe. These populations are exposed to a variety of environments, ranging from warm tropical lowlands to cool high latitude and high altitude regions, in addition to commensal versus wilderness settings (Sprengelmeyer *et al.* 2020). Hence, they are most likely exposed to several unique selective pressures that could be underlying local adaptation and an enrichment of high F_{ST} values.

Alternatively, enrichment for high F_{ST} could also be explained by background selection, which is expected to reduce genetic diversity and therefore result in lower effective population sizes in that genomic region.

Genetic drift is stronger in regions of low N_e, which could increase the differentiation between two populations and produce high F_{ST} (Charlesworth *et al*. 1993). However, a simulation study of background selection targeting stickleback exons found no evidence for background selection increasing F_{ST} outliers (Matthey-Doret & Whitlock 2019).

On the other extreme, the existence of enrichment for low values of F_{ST} suggests that many regions of the genome maintained unexpectedly similar allele frequencies between two populations. Following a population split, neutral evolutionary forces such as genetic drift are expected to increase the genetic differences between two populations. The fact that many regions seemed to have changed less than what was expected due to neutral forces could also be explained by the action of natural selection. This pattern could be the product of shared selective sweeps (i.e. similar selective pressures) taking place in both populations, instead of local adaptation. Shared balancing selection could also be acting at some loci to maintain allele frequencies constant between two populations, perhaps even from before their split time.

We should also acknowledge that the demographic models applied here are simply the best available estimates of population history, and no demographic model fully accounts for the complexity of natural populations. Demographic model misspecification could result in some enrichment of high and/or low F_{ST} values. One potential source of error in

demographic estimation is natural selection. The demographic models were estimated based on tentatively neutral regions of the genome (Sprengelmeyer *et al.* 2020). However, these regions could be under the influence of linked positive and negative selection, with the potential to bias demographic estimation. For example, if the presumed neutral data was substantially affected by either local adaptation or shared sweeps, it could bias the neutral distribution of F_{ST} towards higher or lower values, respectively, making it more difficult to detect F_{ST} outliers in that direction. Nonetheless, previous work suggests that this effect might be weak on demographic inference in *D. melanogaster* (Lange & Pool 2018).

Having hundreds of F_{ST} outlier regions (high or low) between recently diverged population pairs is not unreasonable in light of previous estimates of adaptive divergence. It has been estimated that 19% of substitutions between *D. melanogaster* and *D. simulans* were driven by positive selection (Lange and Pool 2018). Individual genomes from these two species differ at about 5% of sites, although roughly 1% is expected to be driven by segregating polymorphism rather than fixed differences. Given a genome of 120 million bases, this implies an estimated $120{,}000{,}000 \times (0.05 - 0.01) \times 0.19 = 912{,}000$ selectively-driven differences between species. These species are estimated to have diverged about 13,000,000 generations ago (with some uncertainty; Obbard *et al*. 2012), whereas our studied populations are all estimated to have diverged within the past 195,000

generations (Sprengelmeyer *et al*. 2020). Crudely then, we might predict as many as 912,000 × (195,000 / 13,000,000) = 13,680 selectively-driven differences between a population pair such as Ethiopia and Zambia *D. melanogaster*. Hence, although any outlier set may contain both true and false positives for local adaptation, our finding of hundreds of potential targets of adaptation between pairs of *D. melanogaster* populations does not exceed the potentially-expected number of selection-driven differences between them.

In nearly all population pairs, $F_{ST\_FullWin}$ showed a larger enrichment than F_{ST_MaxSNP}. The greater enrichment of $F_{ST\_FullWin}$ persisted when we instead pursued an outlier region removal strategy. In light of the complementary zones of power shown in Figure 1, these results suggest that roughly speaking, there might be a larger contribution of partial hard sweeps than complete soft sweeps to local adaptation among these populations. Furthermore, the importance of partial sweeps in populations of *D. melanogaster* has been proposed previously, including for some of the populations studied here (Pool & Aquadro 2007, Bastide *et al.* 2016, Garud & Petrov 2016, Vy *et al*. 2017). Therefore, seeing fairly low levels of overlap between F_{ST_MaxSNP} and $F_{ST\_FullWin}$ outliers, alongside particularly strong enrichment for $F_{ST\_FullWin}$ outliers, is congruent with the suggested predominance of partial sweeps in the species.

Precautions are needed to ensure high quality $\mathbf{F_{ST\_MaxSNP}}$ ***outliers***

A critically important caveat of using F_{ST_MaxSNP} is that this statistic should be more sensitive to bioinformatic errors than a metric that uses information from all the SNPs in a window. A sequencing or mapping error could cause a single SNP in a window to have a high F_{ST} value, while in a full-window approach such errors are often minimized by being localized to only one or few of the SNPs being aggregated. To reduce false positives from data artifacts, particular consideration should be given to multiple aspects of data preparation and analysis when using F_{ST_MaxSNP}. Prior to population genetic analysis, it is worth considering whether enhanced genotype calling filters are called for, such as increased quality score or depth of coverage thresholds. Excluding sites within a few bp of called indels may also be helpful in reducing erroneous site calls (Lack *et al.* 2015). Furthermore, it is important to ensure that data from all population samples have been collected and assembled the same way. For example, Lange *et al.* (2022) found that a set of SNP-level genetic differentiation outliers from a comparison between individually-sequenced and pool-sequenced population samples were not reliable until genomes from the individually-sequenced population were reassembled using a pipeline analogous to the pool-seq data.

Precautions should also apply to the population genetic analysis itself. Given that F_{ST_MaxSNP} is very sensitive to sample size (Figure 3), variation

in missing data among the sequences of each individual may result in heterogeneous sample sizes for different SNPs in a given window, and therefore using a relatively high minimum sample size threshold for each population is essential. Finally, additional quality control assessment of F_{ST_MaxSNP} outliers following population genetic analysis is desirable. For example, it may be worth confirming that outlier SNPs do not appear to be impacted by depth anomalies suggestive of cryptic structural variation, and are not associated with alignment uncertainty or sub-optimal quality scores. When depth or alignment issues are present, the outlier SNP could potentially be tagging a structural variant under local selection as opposed to representing a pure false positive. In other cases, soft sweeps targeting structural variants might be missed entirely if they fail to strongly alter frequencies at linked SNPs.

The enrichment of nonsynonymous (and UTR) sites among our "top SNPs" in F_{ST_MaxSNP} outlier regions (fig. 8) offers hope that at least in our empirical analysis, many F_{ST_MaxSNP} outlier regions may represent true positives for local adaptation, and that top SNPs may sometimes even reflect causative variants. However, we emphasize that even for true cases of local adaptation, a non-causative SNP may sometimes have a slightly higher F_{ST} value than the causal SNP, simply by chance. And in light of the data quality concerns described above, it makes sense to interpret isolated high F_{ST} SNPs with caution.

Overall then, F_{ST_MaxSNP} outliers may have a wide range of potential significance, ranging from false positives to indicating strong hypotheses for specific variants under selection. Functional experiments may hold particular appeal for F_{ST_MaxSNP} outliers, both to confirm their validity and to investigate the variants they implicate. First, methods such as reciprocal hemizygosity tests (Stern 2014; Turner 2014) may confirm that the implicated genes are associated with detectable trait differences between populations, which would support the outlier F_{ST} signal representing a true positive. Further molecular or transgenic experiments could then assess the consequences of modifying individuals high-F_{ST} variants, to improve our understanding of the precise genetic changes targeted by natural selection.

Summary and future prospects

Here, we have shown that SNP-level F_{ST} (F_{ST_MaxSNP}) offers strong power to detect soft sweeps, and is highly complementary to full-window frequency and haplotype statistics for detecting local adaptation. These results stress the importance of taking into account the different signatures left by different kinds of selective sweeps in the genome when deciding how to perform a genome scan. The raw summary statistics evaluated here can either be applied in parallel, or their signals can be integrated into frameworks such as approximate Bayesian computation and machine

learning. Thus far, the latter methodologies have been used more extensively to detect and classify selective sweeps within a single population (Peter *et al*. 2012, Sheehan & Song 2016, Schrider & Kern 2016, 2017). However, such approaches are equally applicable to the study of local adaptation (Key *et al.* 2014). Future work could investigate whether methods that combine multiple statistics would benefit from including F_{ST_MaxSNP}, potentially increasing their power to detect soft sweeps and their accuracy in classifying different types of sweeps. Because studies of genetic differentiation between populations inherently control for evolutionary variance in the shared ancestral population, local adaptation may offer a better "signal to noise ratio" regarding the types of positive selection acting in natural populations, compared to single population studies. Hence, our results may contribute toward not only an improved ability to detect local adaptation, but also a clearer understanding of adaptation in nature more generally.

Data Availability

No new empirical data were generated for this research. Scripts used in the analyses presented can be found at https://github.com/ribeirots/fst_maxsnp.git.

Supplementary Materials

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We appreciate comments from multiple Pool lab members on this manuscript. We also would like to thank the three anonymous reviewers and associate editor for their comments, which greatly improved the published manuscript. This research was funded by NIH grants R01 GM127480 and R35 GM13630, and by NSF grant DEB 1754745.

References

Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*. 44(1):367–388. doi:10.1146/annurev-ecolsys-110512-135747.

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 437:1149-1152. doi:10.1038/nature04107.

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics*. 4(2):99–110. doi:10.1038/nrg999.

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*. 40:340-345. doi:10.1038/ng.78.

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., & Pool, J. E. (2016). A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics*. 204(3):1307–1319. doi:10.1534/genetics.116.192492.

Catania, F., Kauer, M. O., Daborn, P. J., Yen, J. L., Ffrench-Constant, R. H., & Schlötterer, C. (2004). World-wide survey of an *Accord* insertion and

its association with DDT resistance in *Drosophila melanogaster*. *Molecular Ecology*, 13(8), 2491-2504. https://doi.org/10.1111/j.1365-294X.2004.02263.x.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*. 134(4):1289–1303.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics.* 8(10):e1002905. doi:10.1371/journal.pgen.1002905.

Ewing, G., & Hermisson, J. (2010). MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*. 26(16):2064–2065. doi:10.1093/bioinformatics/btq322.

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*. 18(1):1–16. doi:10.1111/ele.12376.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*. 27(9):489–496. doi:10.1016/j.tree.2012.05.012.

Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PloS Genetics*. 11(2):e1005004. doi:10.1371/journal.pgen.1005004.

Garud, N. R., & Petrov, D. A. (2016). Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics*. 203(2):863–880. doi:10.1534/genetics.115.184002.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics.* 169(4):2335–2352. doi:10.1534/genetics.104.036947.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*. 18(2):337–338. doi:10.1093/bioinformatics/18.2.337.

Jensen, J. D. (2014). On the unfounded enthusiasm for soft selective sweeps. *Nature Communications*. 5(1):5281. doi:10.1038/ncomms6281.

Kaplan, N. L., Hudsonm, R. R., & Langley, C. H. (1989). The "hitchhiking effect" revisited. *Genetics*. 123(4): 887–899. doi:10.1093/genetics/123.4.887.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*. 7(12):1225–1241. doi:10.1111/j.1461-0248.2004.00684.x.

Key, F. M., Peter, B., Dennis, M. Y., Huerta-Sánchez, E., Tang, W., Prokunina-Olsson, L., Nielsen, R., & Andrés, A. M. (2014). Selection on a variant associated with improved viral clearance drives local, adaptive pseudogenization of interferon lambda 4 (*IFNL4*). *PLoS Genetics*, 10(10), e1004681. https://doi.org/10.1371/journal.pgen.1004681.

Kimura, R., Fujimoto, A., Tokunaga, K., & Ohashi, J. (2007). A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS One*. 14;2(3):e286. doi:10.1371/journal.pone.0000286

Lack, J. B, *et al.* 2015. The *Drosophila* Genome Nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*. 199(4):1229–1241. doi:10.1534/genetics.115.174664.

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., Langley, C. H., & Pool, J. E. (2015). The *Drosophila* Genome Nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*, *199*(4), 1229–1241. https://doi.org/10.1534/genetics.115.174664.

Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B., & Pool, J. E. (2016). A thousand fly genomes: an expanded *Drosophila* Genome Nexus. *Molecular Biology and Evolution*. 33:3308-3313. doi:10.1093/molbev/msw195.

Lange, J. D., Bastide, H., Lack, J. B., & Pool, J. E. (2022). A population genomic assessment of three decades of evolution in a natural *Drosophila* population. *Molecular Biology and Evolution*. 39(2). doi:10.1093/molbev/msab368.

Lange, J. D., & Pool, J. E. (2016). A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. *Molecular Ecology*. 25(13):3081–3100. doi:10.1111/mec.13671.

Lange, J. D. & Pool, J. E. (2018). Impacts of recurrent hitchhiking on divergence and demographic inference in *Drosophila*. *Genome Biology and Evolution*. 10(8):1882–1891. doi:10.1093/gbe/evy142.

Lee, K. M., & Coop, G. (2017). Distinguishing among modes of convergent adaptation using population genomic data. *Genetics*. 207(4):1591-1619. doi:10.1534/genetics.117.300417.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. 74(1):175–195. doi:10.1093/genetics/74.1.175.

Matthey-Doret, R., & Whitlock, M. C. (2019). Background selection and F_{ST}: consequences for detecting local adaptation. *Molecular Ecology*. 28(17):3902–3914. doi:10.1111/mec.15197.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*. 23(01):23–35. doi:10.1017/S0016672300014634.

Obbard, D. J., Maclennan, J., Kim, K. W., Rambaut, A., O'Grady, P. M., & Jiggins, F. M. (2012). Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Molecular Biology and Evolution*. 29(11):3459-73. doi:10.1093/molbev/mss150.

Pennings, P. S., & Hermisson, J. (2006a). Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*. 23(5):1076–1084. doi:10.1093/molbev/msj117.

Pennings, P. S., & Hermisson, J. (2006b). Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics.* 2(12):e186. doi:10.1371/journal.pgen.0020186.

Peter, B. M., Huerta-Sanchez, E., & Nielsen, R. (2012). Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genetics*. 8(10):e1003011. doi:10.1371/journal.pgen.1003011.

Pool, J. E., Corbett-Detig, R. B., Sugino, R. P., Stevens, K. A., Cardeno, C. M., Crepeau, M. W., Duchen, P., Emerson, J. J., Saelao, P., Begun, D. J., & Langley, C. H. (2012). Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genetics*, 8(12), e1003080. https://doi.org/10.1371/journal.pgen.1003080.

Pool, J. E., & Aquadro, C. F. (2007). The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Molecular Ecology*. 16(14):2844–2851. doi:10.1111/j.1365-294X.2007.03324.x.

Pool, J. E., Braun, D. T., & Lack, J. B. (2017). Parallel evolution of cold tolerance within *Drosophila melanogaster*. *Molecular Biology and Evolution*. 34(2):349–360. doi:10.1093/molbev/msw232.

Ralph, P. L., & Coop, G. (2015). The role of standing variation in geographic convergent adaptation. *American Naturalist*. 186(S1):S5-S23. doi: 10.1086/682948.

Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F., & Carroll, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science*. 326(5960):1663–1667. doi:10.1126/science.1178357.

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*. 105(3):767–779. doi:10.1093/genetics/105.3.767.

Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences*. 104(Suppl 1):8641–8648. doi:10.1073/pnas.0700643104.

Sakamoto, T., & Innan, H. (2019). The evolutionary dynamics of a genetic barrier to gene flow: From the establishment to the emergence of a peak of divergence. *Genetics*. 212(4):1383-1398. doi:10.1534/genetics.119.302311

Santiago, E., & Caballero, A. (2005). Variation after a selective sweep in a subdivided population. *Genetics*. 169(1):475-483. doi:10.1534/genetics.104.032813

Schlenke, T. A., & Begun, D. J. (2004). Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proceedings of the National Academy of Sciences*. 101(6):1626–1631. doi:10.1073/pnas.0303793101.

Schrider, D. R., & Kern, A. D. (2016). S/HIC: Robust identification of soft and hard sweeps using machine learning. *PLoS Genetics*. 12(3):e1005928. doi:10.1371/journal.pgen.1005928.

Schrider, D. R., & Kern, A. D. (2017). Soft sweeps are the dominant mode of adaptation in the human genome. *Molecular Biology and Evolution*. 34(8):1863–1877. doi:10.1093/molbev/msx154.

Sheehan, S., & Song, Y. S. (2016). Deep learning for population genetic inference. *PLoS Computational Biology*. 12(3):e1004845. doi:10.1371/journal.pcbi.1004845.

Sprengelmeyer, Q. D., Mansourian, S., Lange, J. D., Matute, D. R., Cooper, B. S., Jirle, E. V., Stensmyr, M. C., & Pool, J. E. (2020). Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Molecular Biology and Evolution*, 37(3), 627–638. https://doi.org/10.1093/molbev/msz271.

Stern, D. L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends in Genetics*, 30(12), 547-554. https://doi.org/10.1016/j.tig.2014.09.006.

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*. 25(10):2144–2164. doi:10.1111/mec.13606.

Turner, T. L. (2014). Fine-mapping natural alleles: quantitative complementation to the rescue. *Molecular Ecology*, 23(10), 2377-2382. https://doi.org/10.1111/mec.12719.

Vy, H. M. T., Won, Y. J., & Kim, Y. (2017). Multiple modes of positive selection shaping the patterns of incomplete selective sweeps over African populations of *Drosophila melanogaster*. *Molecular Biology and Evolution*. 34(11):2792–2807. doi:10.1093/molbev/msx207.

Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M., & Hill, W. G. (2005). Measures of human population structure show heterogeneity among genomic regions. *Genome Research*. 15(11):1468–1476. doi:10.1101/gr.4398405.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *American Naturalist*. 186(S1):S74–S89. doi:10.1086/682405.

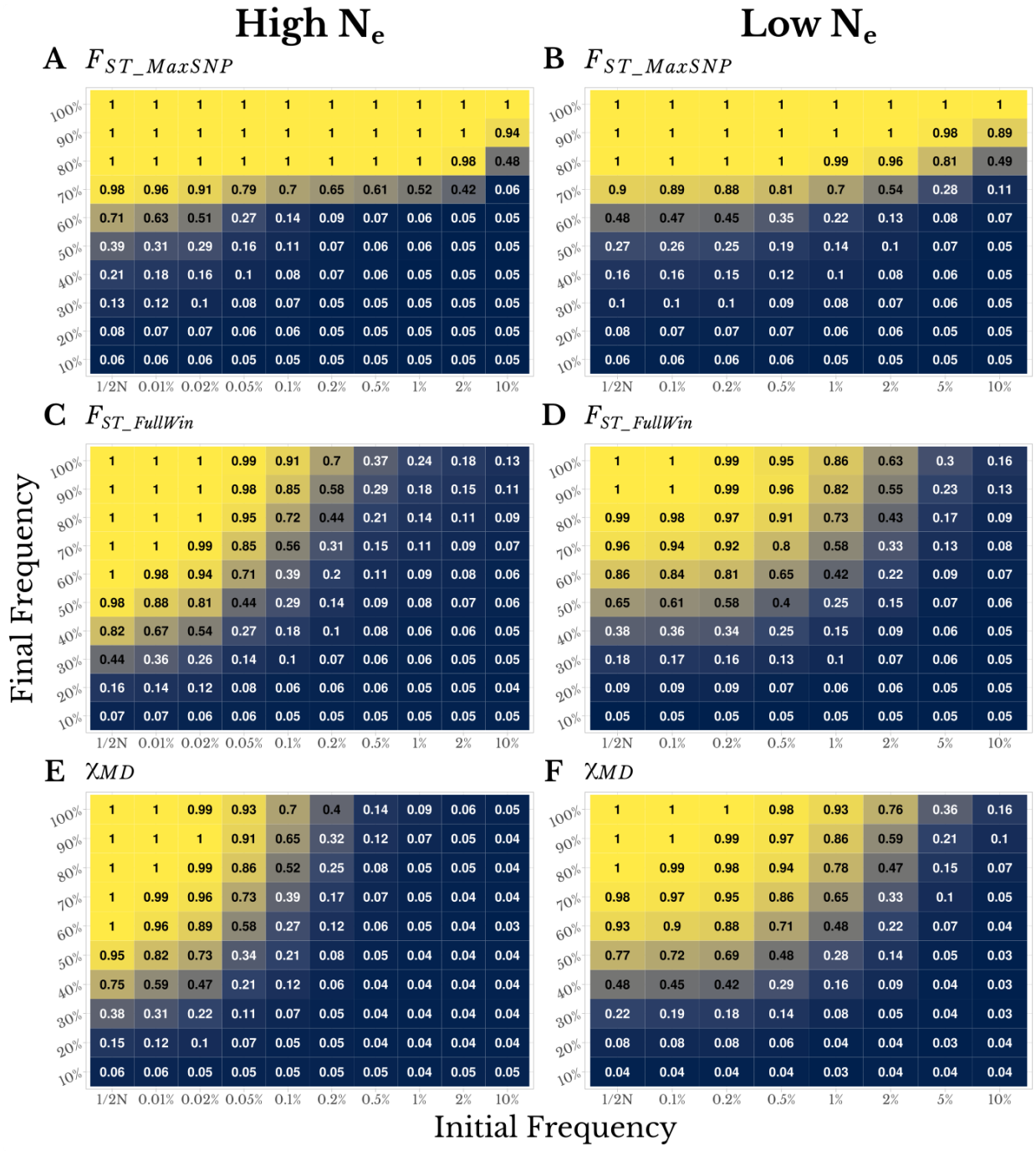


Figure 1. SNP-level F_{ST} and full-window statistics show complementary power to detect local adaptation, depending on the type of selective sweep simulated. Numbers and colors in each panel both depict statistical power to detect local adaptation, in high N_e populations (s=0.001, left column) and

low N_e populations (s=0.01, right column). In each panel, the x-axis illustrates the pre-selection frequency of a favored variant (with the left column indicating selection on newly-occurring mutations) and the y-axis illustrates the final frequency of the sweep (with the top row showing complete sweeps). Detection power is shown for (A and D) F_{ST_MaxSNP}, (B and E) $F_{ST\_FullWin}$, and (C and F) χ_{MD}. These results are based on a demographic history of simple isolation between two populations without change in population size, with a split time of $0.2N_e$ generations.

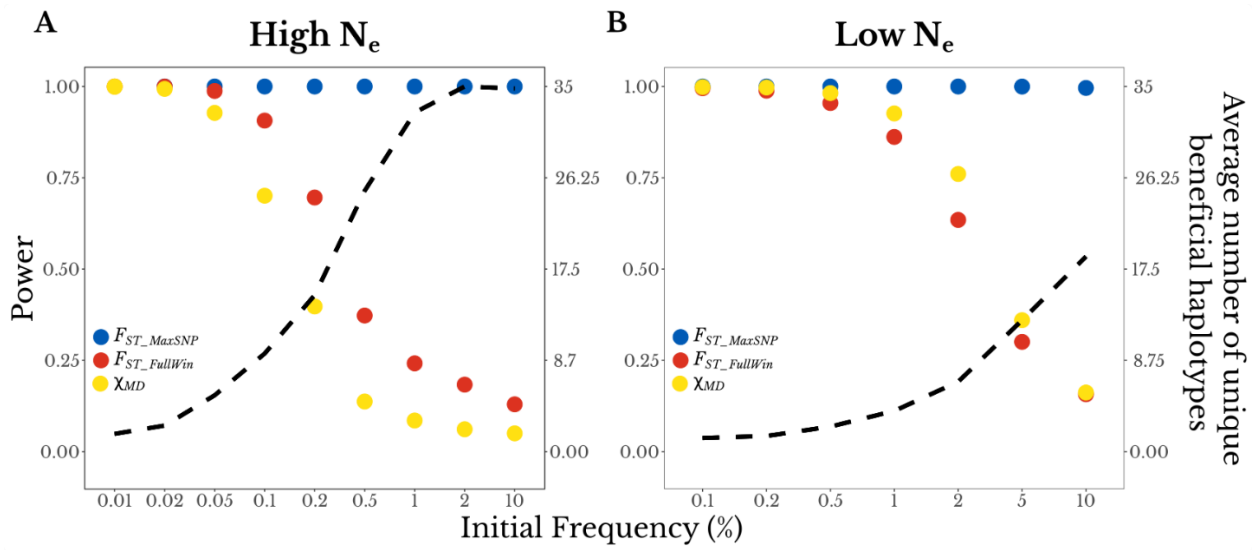


Figure 2. F_{ST_MaxSNP} shows an increasing power advantage as sweeps become softer. For complete sweeps with a range of initial frequencies (x-axis), the two y-axes show detection power for each statistic (left axis, dots) and the average number of unique beneficial haplotypes present at the end of the simulation (right axis, dashed line). Results are shown for (A) high N_e populations (s = 0.001) and (B) low N_e populations (s = 0.01), for the same demographic history as in Figure 1.

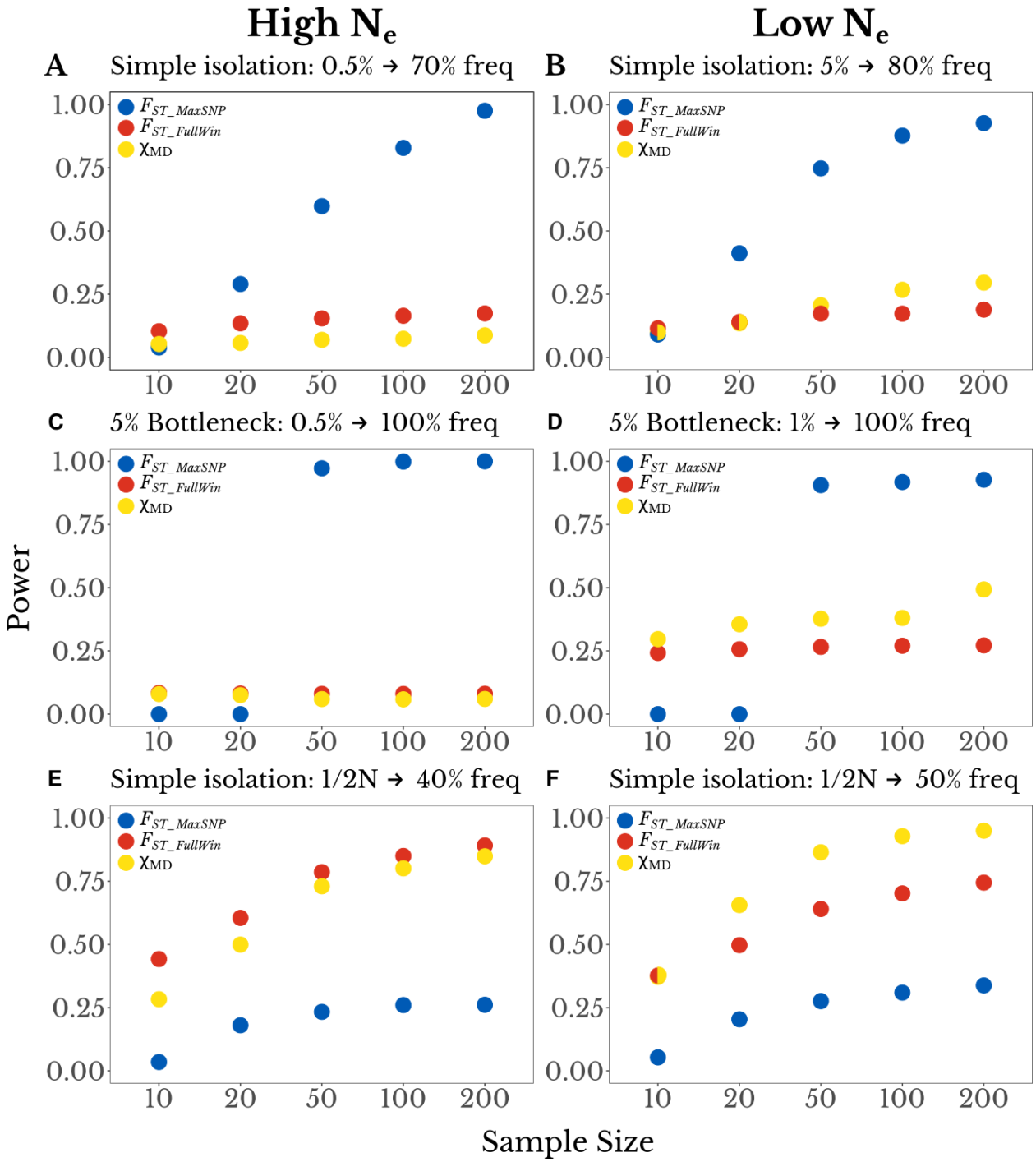


Figure 3. The power of F_{ST_MaxSNP} is particularly sensitive to sample size. Here, the power of each statistic (y-axis) is plotted as a function of sample size (x-axis; number of chromosomes per population). We found that depending on sample size, F_{ST_MaxSNP} outperforms $F_{ST\_FullWin}$ and χ_{MD} for a simple

isolation model, for: (A) a high N_e population with initial beneficial allele frequency of 0.005 and final frequency of 0.70,and (B) a low N_e population with initial frequency 0.05 and final frequency of 0.80. Similar results were observed for a complete soft sweep with a population bottleneck of 0.05, except that the loss of power for F_{ST_MaxSNP} was more immediate at lower sample sizes, for: (C) a high N_e population with initial frequency 0.05, (D) a low N_e population with initial frequency 0.01. For partial hard sweep scenarios where $F_{ST\_FullWin}$ and χ_{MD} outperform F_{ST_MaxSNP}, all three statistics show more gradual sample size effects, specifically for new mutations and: (E) a final frequency of 0.40 in a high N_e population, and (F) a final frequency of 0.50 in a low N_e population.

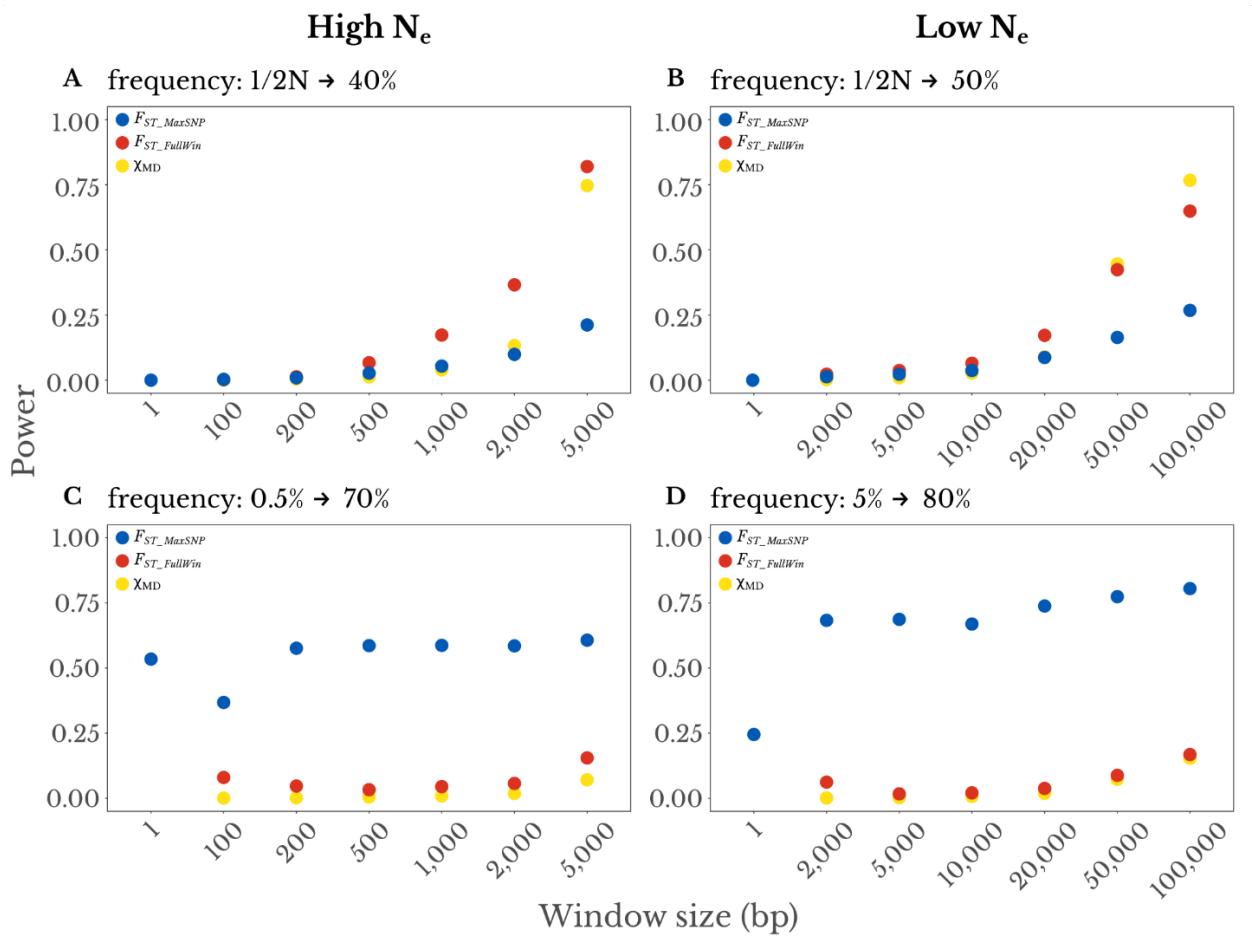


Figure 4. Varying window size does not reveal a single statistic with broad detection power. The top panels show partial hard sweeps for which $F_{ST\_FullWin}$ and χ_{MD} outperform F_{ST_MaxSNP}: (A) a high N_e population with a final beneficial allele frequency of 0.40, And (B) a low N_e population with a final frequency of 0.50. The bottom panels show mostly complete soft sweeps for which F_{ST_MaxSNP} outperforms $F_{ST\_FullWin}$ and χ_{MD}: (C) a high N_e population with an initial beneficial allele frequency of 0.005 and final frequency of 0.70, and (D) a low N_e population with initial frequency 0.05 and final frequency 0.80. These power values reflect a Bonferroni-corrected significance threshold

to reflect the relatively larger number of smaller windows needed. Results do not suggest that any statistic in a smaller window size captures the advantages of both F_{ST_MaxSNP} and the full-window statistics.

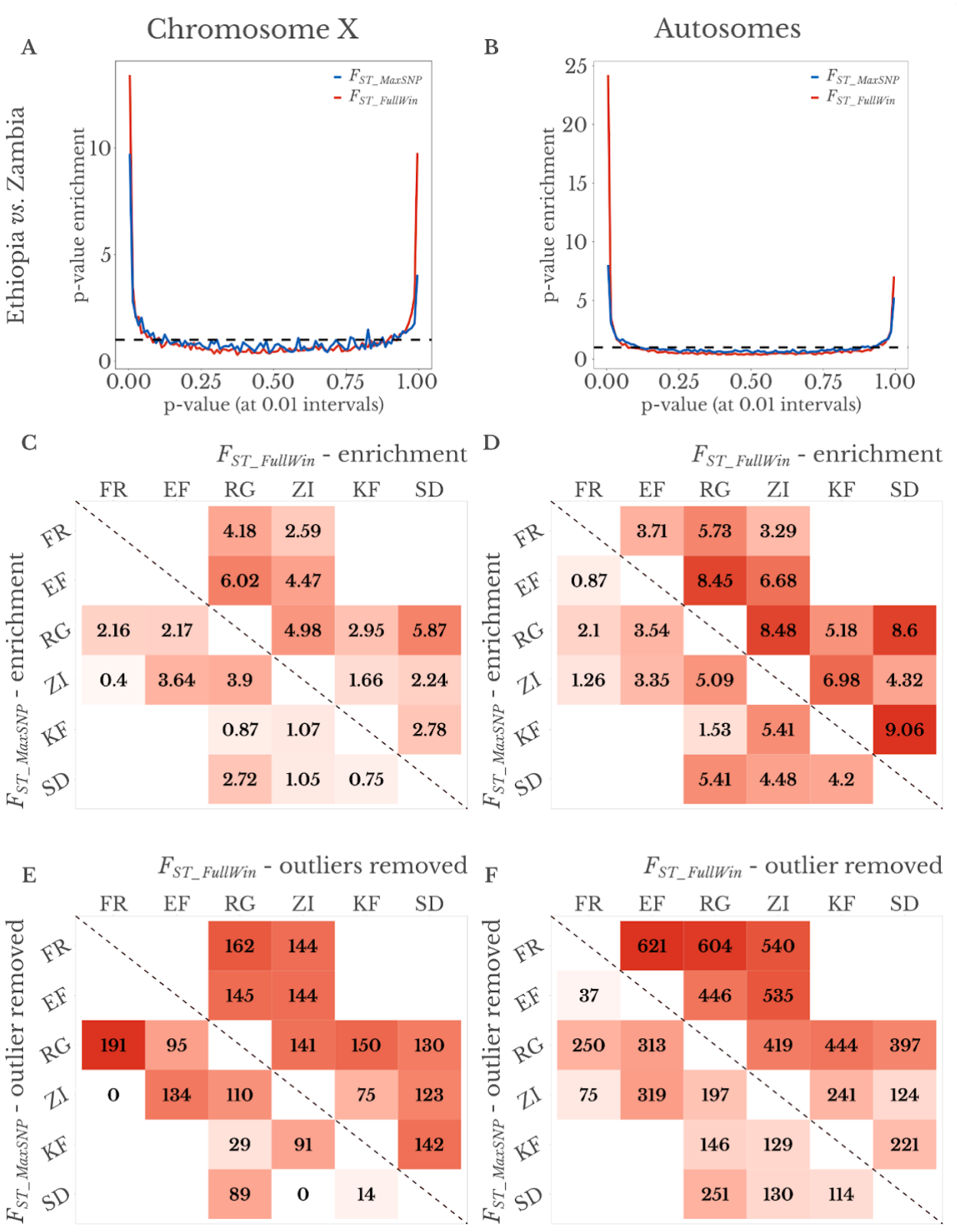

A
Chromosome X
Ethiopia vs. Zambia
F_{ST_MaxSNP}
$F_{ST\_FullWin}$
p-value enrichment
p-value (at 0.01 intervals)
B
Autosomes
C
$F_{ST\_FullWin}$ - enrichment
F_{ST_MaxSNP} - enrichment
FR EF RG ZI KF SD
4.18 2.59
6.02 4.47
2.16 2.17 4.98 2.95 5.87
0.4 3.64 3.9 1.66 2.24
0.87 1.07 2.78
2.72 1.05 0.75
D
$F_{ST\_FullWin}$ - enrichment
F_{ST_MaxSNP} - enrichment
3.71 5.73 3.29
0.87 8.45 6.68
2.1 3.54 8.48 5.18 8.6
1.26 3.35 5.09 6.98 4.32
1.53 5.41 9.06
5.41 4.48 4.2
E
$F_{ST\_FullWin}$ - outliers removed
F_{ST_MaxSNP} - outlier removed
162 144
145 144
191 95 141 150 130
0 134 110 75 123
29 91 142
89 0 14
F
$F_{ST\_FullWin}$ - outlier removed
F_{ST_MaxSNP} - outlier removed
621 604 540
37 446 535
250 313 419 444 397
75 319 197 241 124
146 129 221
251 130 114

Figure 5. F_{ST_MaxSNP} and $F_{ST\_FullWin}$ both show outlier enrichment between natural populations of *D. melanogaster*. (A and B) Ethiopia-Zambia F_{ST_MaxSNP} and $F_{ST\_FullWin}$ values on (A) chromosome X and (B) autosomes show enrichment of low (right) and especially high values (left), based on the distribution of p-values obtained from neutral demographic simulations. (C and D) F_{ST_MaxSNP} (lower diagonal) and $F_{ST\_FullWin}$ (upper diagonal) both show enrichment of high outliers on (C) chromosome X and (D) combined autosome arms. $F_{ST\_FullWin}$ shows a greater enrichment in nearly all cases. (E and F) Number of outlier regions that were removed to erase the signature of enrichment for high F_{ST_MaxSNP} (lower diagonal) and $F_{ST\_FullWin}$ (upper diagonal) for each population on (E) chromosome X and (F) the combined autosome arms. $F_{ST\_FullWin}$ was associated with a greater outlier region enrichment for most population pairs, reinforcing the window-level patterns shown in (C) and (D). Populations: SD = South Africa. ZI = Zambia. KF = Kafue, Zambia. RG = Rwanda. EF = Ethiopia. Population pairs that were not present in the same demographic model were not evaluated. Color scale ranges from the minimum to maximum value within each panel.

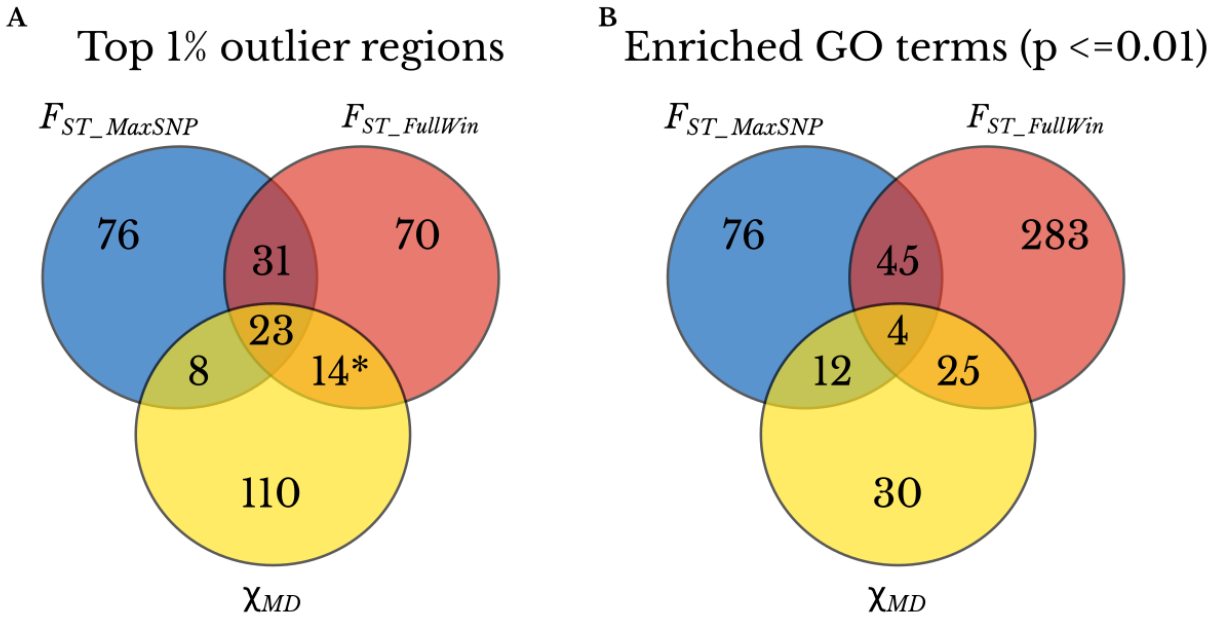
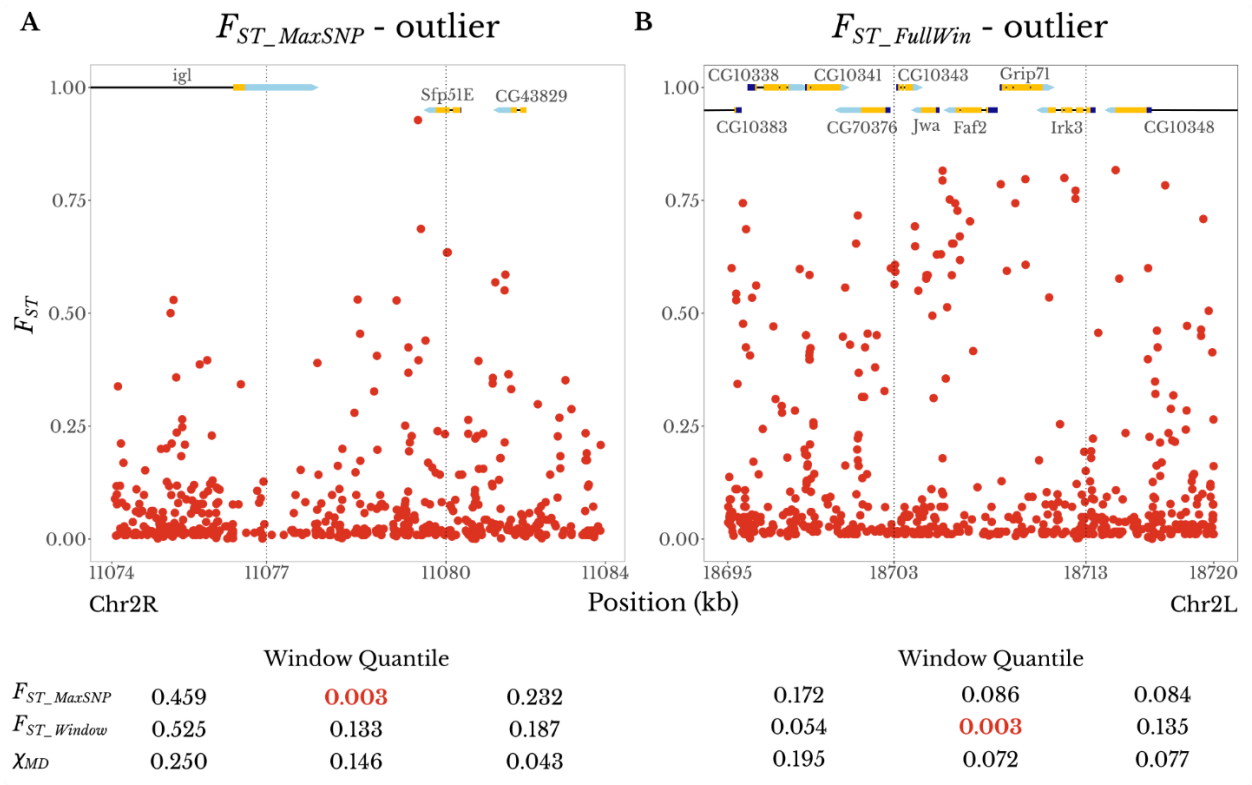


Figure 6. The three statistics detect mostly unique genomic regions and functional categories. (A) Overlap between the top 1% outlier regions detected with F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD}. * indicates the average number of outlier regions between the two statistics: 15 $F_{ST\_FullWin}$ outlier regions exclusively overlap χ_{MD} outliers and 13 χ_{MD} outlier regions exclusively overlap $F_{ST\_FullWin}$ outliers. (B) Overlap between enriched GO terms with raw p-value <= 0.01, out of a total of 47,496 GO terms, based on the outlier regions detected with F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD}.



	Window Quantile			Window Quantile		
F_{ST_MaxSNP}	0.459	0.003	0.232	0.172	0.086	0.084
F_{ST_Window}	0.525	0.133	0.187	0.054	0.003	0.135
χ_{MD}	0.250	0.146	0.043	0.195	0.072	0.077

Figure 7. Examples of the distinct SNP-level F_{ST} landscapes associated with F_{ST_MaxSNP} versus $F_{ST\_FullWin}$ outliers. Each plot shows an outlier window for an Ethiopia-Zambia F_{ST} statistic, plus its adjacent windows. Dashed vertical lines delimit the boundaries of the windows. Numbers under each window are the empirical quantiles of that window's statistic (F_{ST_MaxSNP}, $F_{ST\_FullWin}$, and χ_{MD}) in relation to the chromosome arm-wide distribution of the same statistic, with the outlier (quantile < 0.01) value in red. (A) An outlier window for F_{ST_MaxSNP} (center) shows a peak-like F_{ST} landscape with one particularly differentiated SNP. (B) An outlier window for $F_{ST\_FullWin}$ (center) shows a broad plateau of fairly high F_{ST} values. Gene names and structures are shown at

the top of each plot. Protein-coding exons are in yellow, while 5' and 3' untranslated regions are in dark blue and light blue, respectively.

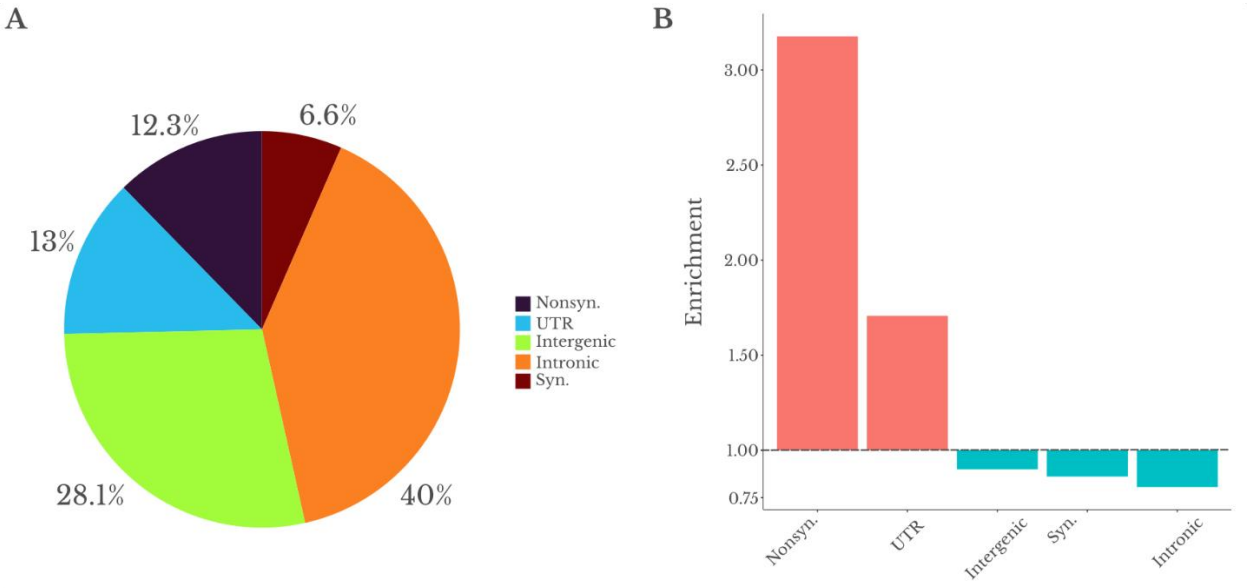


Figure 8. The most differentiated SNPs in F_{ST_MaxSNP} top outlier regions are strongly enriched for site categories known to experience more frequent selection. (A) Proportional distribution of these top SNP among five different classes: nonsynonymous, untranslated regions (UTR), intergenic, synonymous, and intronic. (B) Enrichment analysis of each the five classes in the outlier regions for F_{ST_MaxSNP} in comparison to genome-wide distribution for all SNPs with similar minor allele frequencies.

Chapter 3: Recombinant Inbred Line panels inform the genetic architecture and interactions of adaptive traits in *Drosophila melanogaster*

Co-author contribution acknowledgment

We would like to thank Quentin D. Sprengelmeyer, Yuheng Huang, Derek M. Benson, Megan S. Orr, and Zachary C. Johnson for their contributions to the collection of phenotypic data used in this project. We also would like to thank Matthew J. Lollar and Russell B. Corbett-Detig for their contributions to DNA library preparation and DNA sequencing, as well as in implementing the code and pipeline used in the data analysis of this project.

Abstract

The distribution of allelic effects on traits, along with their gene-by-gene and gene-by-environment interactions, determines the phenotypes available for selection and the trajectories of adaptive variants. Nonetheless, uncertainty persists regarding the effect sizes underlying adaptations and the importance of genetic interactions. Herein, we aimed to investigate the genetic architecture and the epistatic and environmental interactions involving loci that contribute to multiple adaptive traits using two new panels of *Drosophila melanogaster* recombinant inbred lines (RILs). To better fit our data, we re-implemented functions from R/qtl (Broman *et al*. 2003) using additive genetic models. We found 14 quantitative trait loci

(QTL) underlying melanism, wing size, song pattern, and ethanol resistance. By combining our mapping results with population genetic statistics, we identified potential new genes related to these traits (e.g. *Vha68* and *shakB* for wing length). None of the detected QTLs showed evidence of epistasis, and our power analysis supports that we should have seen at least one significant interaction if sign epistasis or strong positive epistasis played a pervasive role on trait evolution. In contrast, we did find roles for gene-by-environment interactions involving pigmentation traits. Overall, our data suggest that the genetic architecture of adaptive traits often involves alleles of detectable effect, that strong epistasis does not always play a role in adaptation, and that environmental interactions can modulate the effect size of adaptive alleles.

Introduction

Adaptive evolution is the result of natural selection, which acts by increasing the frequency of beneficial traits in a population. A number of important uncertainties persist regarding the genetic architectures of adaptive trait changes in nature, including the number and effect sizes of contributing variants, and their frequencies before and after the action of positive selection (Pritchard *et al*. 2010, Savoleinen *et al*. 2013, Barghi *et al*. 2020). Furthermore, research on the genetic basis of adaptation has largely been centered on the individual additive effects of each gene underlying

the adaptive trait, whereas the role of gene-by-gene and gene-by-environment interactions in adaptation has remained more elusive (Whitlock *et al*. 1995, Malmberg & Mauricio 2005, Martin & Lenormand 2006, Bank 2022).

Theoretical and empirical studies have yielded varied findings concerning the number and strength of loci contributing to adaptive trait change. Fisher (1930) argued that traits were governed by numerous genetic variants of very small magnitude, and that natural selection proceeded via minute incremental changes in allele frequency at these loci. Extending Fisher's geometric model, Orr (1998) found that in a sequential model of genetic adaptation, a few larger effect sizes should tend to be followed by many smaller ones, yielding an exponential distribution of effects. However, adaptation in natural populations may sometimes depart from the assumptions of such models in multiple respects. First, the role of standing variation (as opposed to new mutations) in the adaptive process continues to be investigated (Fuhrmann *et al*. 2023, Schlötterer 2023), and these two sources of genetic variation may differ in their effect sizes, dominance, and other properties (e.g. Orr & Betancourt 2001). Second, the genetic architecture of local adaptation, in which populations in a heterogeneous landscape adapt to local conditions in the presence of migration, has a stronger bias towards alleles of large effect sizes (Griswold 2006, Yeaman & Whitlock 2011, Yeaman & Otto 2011). Third, genetic

interactions may alter the dynamics of adaptive variants, as further discussed below.

In regards to the role of epistasis in adaptation, a discussion has existed since the onset of the Modern Synthesis, with some arguing that it is largely irrelevant to the trajectory of Darwinian selection (Fisher 1930, Cohan *et al*. 1989, Hill *et al*. 2008, Crow 2010) and others proposing that it plays an important role (Fenster *et al*. 1997, Hansen 2013, Wright 1931). On the one hand, epistasis might not be expected to strongly affect adaptation because alleles at unlinked loci can segregate independently and selection would act on the combined effect of each allele against different genetic backgrounds. Nonetheless, epistasis has the potential to alter the rate at which adaptive trait changes can occur by producing larger or smaller effects in some of these backgrounds, and in certain contexts, it may wield a stronger influence on the course of adaptation.

Epistasis has been shown to play a role in many evolutionary processes, including the evolution of sex and recombination, speciation, canalization, maintenance of genetic variation, and the evolution of co-adapted gene complexes (Hansen 2013). How epistatic interactions alter the response to directional selection is determined by the nature of the epistatic interaction (Hansen 2013). Some kinds of epistasis affect the magnitude of the phenotypic effects, such as positive and negative epistasis, and another kind, called sign epistasis, changes the direction of

the effect. For positive (synergistic) epistasis, in which the combined effect of both alleles is greater than the sum of the individual effects, additive genetic variance would increase, and the mean fitness in an adapting population is expected to change faster and reach a higher plateau than without epistasis. For negative epistasis, in which the combined effect is lower than the sum of the individual effects but still qualitatively consistent with the additive effects, additive genetic variance decreases and the mean fitness during adaptation changes more slowly and reaches a lower plateau than without epistasis. Particularly in the context of adaptation, negative epistasis is also referred to as diminishing returns epistasis. For sign epistasis, whether the allele will contribute to a larger or smaller phenotype will depend on the allele at another locus. In this case, the evolutionary dynamics can be more complex, and the fate of one allele is attached to the other, but while both variants persist, their interaction may impede adaptation. Furthermore, to the extent that epistasis masks the phenotypic influence of some variants, it may facilitate the accumulation of cryptic genetic variation for a given trait, which may subsequently be exposed to selection when elements of the genetic background (or the environment) change (Gibson & Dworkin 2004, Steiner *et al*. 2007).

Despite the pervasiveness of epistasis at the genomewide scale (Huang *et al*. 2012), as well as between adaptive variants within proteins (Lunzer *et al*. 2010, Gong & Bloom 2014, Starr & Thorthon 2016), the

prevalence of epistatic interactions among adaptive variants that contribute to the evolution of more complex adaptive traits remains incompletely understood. Studies of crop domestication have reported examples of epistasis among artificially selected alleles, including grapevine varieties (Duchêne *et al*. 2012) and many examples in maize domestication (Stitzer & Ross-Ibarra 2018). Still, compared to wild alleles, domesticated alleles were usually less sensitive to the genetic background (reviewed in Doust *et al*. 2014). Epistasis was also found among adaptive quantitative trait loci (QTLs) underlying flowering time in *Arabidopsis thaliana* (Juenger *et al*. 2002). In a recent genome-wide screen in yeast, using genes known to be involved in a high number of interactions, 24% of the adaptive variants were strain-specific and indicative of epistasis (Ang *et al*. 2023). Positive epistasis has been shown to underlie the evolution of bacteria with multiple antibiotic resistance alleles (Trindade *et al*. 2009). Examples in animals involve epistasis between alleles underlying adaptive coloration in butterflies (Papa *et al*. 2013), negative epistasis for traits underlying fitness in populations of *Drosophila melanogaster* experimentally selected on a chronic larval malnutrition regimen (Vijendravarma & Kawecki 2013), and epistasis between genes underlying coat color in oldfield mice, in which the effect of one allele is only expressed in the presence of another allele (Steiner *et al*. 2007).

The genetic architecture of adaptive traits can also be affected by gene-by-environment interactions because the effect of a given allele can change depending on its environment (Zeng *et al*. 1999, Mackay 2001). Examples of interactions between the environment and adaptive alleles include *D. melanogaster* reproductive performance at different temperatures (Fry *et al*. 1998), drought stress adaptation in wheat (Mathews *et al*. 2008), and local adaptation reflected in biomass for switchgrass (Lowry *et al*. 2019). Similarly to epistatic interaction, gene-by-environment interaction can also uncover crypt genetic variation (Gibson & Dworkin 2004). The chaperone protein Hsp90 offers an example of how genetic and environmental interactions can uncover cryptic variation underlying discrete and continuous traits (Rutherford & Lindquist 1998, McGuigan & Sgro 2009, Flynn *et al*. 2020).

Here, we contribute to the understanding of the genetic architecture of adaptive traits and the emerging knowledge about genetic and environmental interactions by identifying adaptive quantitative trait loci (QTLs) underlying recent local adaptation in natural populations of *D. melanogaster* and asking whether there is evidence of epistasis among them and, for a subset, whether they interact with the environment. Using two newly-described panels of recombinant inbred lines (RILs), we investigate several traits that appear to have evolved directly or indirectly under adaptive differentiation between an ancestral range population (Zambia)

and a population from either France or highland Ethiopia, regions colonized by *D. melanogaster* approximately 2 kya (Sprengelmeyer *et al*. 2020).

The traits examined here (cold tolerance, ethanol resistance, pigmentation, song pattern, and wing length) show evidence of being either direct targets of local adaptation or else pleiotropic readouts of local adaptation targeting correlated traits. The pigmentation traits show strong correlations with environmental variables, especially ultraviolet radiation (Bastide *et al*. 2014), which reaches particularly high levels in the Ethiopian highlands. Larger wings may help Ethiopian flies navigate in cooler, thinner air, and wing size was found to show unusually strong differentiation (Q_{ST} = 0.985) between the Ethiopia and Zambia populations, compared to genome-wide genetic differentiation (Lack *et al*. 2016), indicating a role for selection in this trait's differentiation. France and Zambia have markedly different ethanol resistance, with a Q_{ST} of 0.548 exceeding patterns of genome-wide genetic differentiation (Sprengelmeyer *et al*. 2021). Cold developmental survival shows a similarly consistent differentiation between France and Zambia in particular (Huang *et al*. 2021), with clear presumptive adaptive value in light of climate differences between these regions. A male song trait – the proportion of pulses classified as slow (Clemens *et al*. 2018) – was included based on strong population differentiation (Lollar *et al*. *in Prep*), with a similarly extreme Q_{ST}

of 0.797. Although directional selection has not been previously reported to have acted directly on song traits within *D. melanogaster*, it is also possible that this trait difference reflects selection on a pleiotropically connected trait, such as synaptic function in a novel thermal environment (Pool *et al*. 2017).

For each of the above traits, we first identify QTLs and candidate genes that may underlie each of these traits. Then, we quantify the strength of evidence for epistasis impacting each detected QTL, and then assess the overall signal of epistasis across traits. Lastly, we investigate the reaction norm of the QTLs underlying a subset of the traits, related to pigmentation, that were studied in two different temperature treatments. These investigations provide new insights into both the genetic basis of adaptive trait changes and the dependence of adaptive alleles on genetic interactions.

Methods

Mapping cross design

We report two new Recombinant Inbred Line (RIL) mapping panels of *Drosophila melanogaster*. Each RIL set was derived from a cross between two inbred lines from distinct geographic populations, in each case pairing a strain from the southern-central African ancestral range of the species (Sprengelmeyer *et al*. 2020) with a strain from a cooler derived

environment. One cross was between a Zambian inbred line (ZI215N, sequenced in Chapter 4 of this dissertation) and a highland Ethiopian inbred line (EF43N, sequenced in Lack *et al*. 2016); the other was between a different Zambian inbred line (ZI418N, sequenced in Sprengelmeyer & Pool 2021) and a French inbred line (FR320N, sequenced in Lack *et al*. 2016). Each cross was allowed to interbreed in an intercross design for 13 and 12 non-overlapping generations, in the Ethiopian and French cross, respectively. Then, the offspring of individual females were inbred for 5 generations to create each RIL. We obtained 293 and 328 RILs for the Ethiopian and French panels, respectively. The Zambian lines were collected in Siavonga, the Ethiopian line was collected in Fiche, and the French line was collected in Lyon (Lack *et al*. 2015).

Phenotypes

Ethiopian flies have increased wing area, even after correcting size relative to their larger body mass (Lack *et al*. 2016). We measured the wing length of female flies that were at least 3-days old. 10 female flies per RIL were photographed using an Amscope SM-4TZZ-144A dissection microscope under CO_2 anesthesia. The pictures were scored with ImageJ (Abramoff *et al*. 2004). Wing length was measured as the distance from the intersection of the L4 longitudinal vein and the anterior cross vein to the L3 longitudinal vein intersection with the wing margin.

Ethiopian flies are darker than Zambian flies (Bastide *et al*. 2014). We photographed and measured two to five females per RIL using a dissection scope. Resolution (3584 x 2748 pixels), gain (3.0), exposure time (99.84 ms), magnification (20x), and illumination level were kept constant using an Amscope adaptor for LED lamp at maximum lighting. White balance was set at the same resolution using ColorChecker white balance. We measured three common traits for pigmentation: the pigmentation of the mesopleuron in greyscale proportion ranging from 0 for white to 1 for black (herein, mesopleuron, a thorax trait), the pigmentation of the background of the fourth abdominal segment in the same greyscale proportion (herein, A4 Background), and the proportion of the fourth abdominal segment that was covered by the black stripe (herein, stripe ratio). Relatively low correlations among these three traits were observed among independent isofemale strains within the Ethiopia and Zambia populations ($r < 0.35$), except for a higher correlation between mesopleural and A4 Background for Ethiopia specifically ($r = 0.646$; Bastide *et al*. 2014).

Pigmentation is also plastically influenced by temperature, and the Ethiopian population is located in a colder region, on a high plateau 3 km above sea level. Therefore, we measured the three pigmentation traits in flies raised at two temperatures (25 °C and 15 °C), to uncover potential cryptic variation that could be present in the warmer Zambian (midpoint temperature: 25 °C) but is only expressed in the colder Ethiopian (midpoint

temperature: 11 °C) population. We also report plasticity for these three pigmentation traits, as the difference in the trait measured at 25 °C and 15 °C.

French flies have higher resistance to ethanol (Cohan & Graf 1985, Sprengelmeyer & Pool 2021) and cold (Pool *et al*. 2017, Huang *et al*. 2021) than Zambian flies. Ethanol resistance was measured as the average time 300 female flies, 3- to 5-day old, from each RIL needed to become immobile when exposed to 18% ethanol. The flies were placed in 50 ml falcon tubes with ethanol-saturated tissue placed on the bottom, and the number of mobile flies was scored every 15 minutes for six hours. Cold tolerance was measured as egg to adult survival when raised at 15 °C (Huang *et al*. 2021).

The song pattern trait was the ratio of slow pulses over all pulses (slow and fast) in the song made by males during courtship. Although not expected to be under local environmental selection, male courtship is often under strong sexual selection. In addition, a large number of nervous system genes appear to have evolved between warm- and cold-adapted *D. melanogaster* populations (Pool *et al*. 2017), representing one potential source of pleiotropic effects on song traits. We found that Zambian male flies display a greater fraction of slow to fast pulses than French flies, a novel pulse mode classification recently discovered within the *D. melanogaster* species complex (Clemens *et al*. 2018). The song data was recorded in the

Stern Lab at Janelia Farms Research Campus, using materials and methods described by Arthur and colleagues (Arthur *et al*. 2013). Song was annotated without human intervention with SongExplorer (Arthur *et al*. 2021). Annotations were classified with locally modified versions of BatchSongAnalysis (https://github.com/dstern/BatchSongAnalysis), using previously trained *D. melanogaster* models provided by Ben Arthur (Lollar *et al*. *in Prep*).

We expected that there might be some level of correlation among the traits of any given cross purely due to either linkage among causative variants or else pleiotropy. We calculated the pairwise correlation between all the traits measured for each cross to investigate the degree to which they are correlated.

DNA Extraction and Sequencing

For each RIL, we extracted DNA from five to ten female flies from each line. The flies were crushed and homogenized using pipette tips in strip tubes, in a solution of 100 μl of lysis buffer (50 mM Tris-HCl with pH = 8, 100 mM EDTA, 100 mM NaCl, and 0.5% SDS) and 2 μl of proteinase K (IBI Scientific). The samples were incubated to inactivate the proteinase K (55 °C for 5 minutes, 95 °C for 10 minutes, and 25 °C hold). The supernatant of the fly lysate was transferred to new tubes, and we added SPRI beads (Sera-mag SpeedBeads SPRI beads, [GE Healthcare

24152105050250]) in a 1.8 to 1 proportion of beads to lysate. The beads were used in a solution of PEG 8000 (9 g), Tween 20 (27.5 μl), 5M NaCl, and 3 mL TE Buffer (composed of 100 μl of 1 mM EDTA, 500 μL of 10 mM Tris HCl). 1 mL of the beads were added to the buffer mix and filled with water until the final volume was mL. The tubes were incubated for at least three minutes at room temperature for DNA to bind to the beads and then placed on a magnetic plate for five minutes to remove the beads from the solution. The supernatant was discarded, and the beads were washed with two rounds of 80% ethanol (90 μL per well). DNA was eluted beads with 30 μL of H_2O and then used the magnetic plate to transfer only the resuspended DNA to new tubes.

Genomic libraries were prepared following Adams *et al.* (2020), with size selection, cutoff at 300 bp, and cleanup performed using the Zymo Select-a-Size DNA clean & Concentrator [catalog No. D4080]. Libraries were sequenced at the UW-Madison Biotechnology Center on the Illumina NovaSeq 6000 platform, with pair-ended reads of 150 bp.

DNA Alignment and Ancestry Calling

The reads were mapped to the *D. melanogaster* (v5.57) reference genome using BWA-MEM, version 0.7.17 (Li 2013). We used Ancestry HMM (Corbett-Detig & Nielsen 2017) to estimate ancestry along the genome of each RIL, summarized in windows defined to contain 1,000

SNPs in the Zambian population (~19kb average size). We used a recombination map based on Comeron *et al*. (2012) but modeled only one generation of recombination when using Ancestry HMM, in order to focus on conservatively-defined ancestry switches (pipeline available on https://github.com/ribeirots/RILepi).

QTL Mapping

To test whether adaptive loci showed evidence of epistatic interactions we first used QTL mapping to identify the adaptive loci and then performed epistasis tests between the focal loci and all other loci in the genome. Each locus is a genomic window of ~19kb, defined and genotyped as described above.

To identify loci underlying adaptive trait changes, we reimplemented the R/qtl (Broman *et al*. 2003) *scanone()* function in an additive framework using Python3 (Van Rossum & Drake 2009), modifying the genotypes to be coded as numeric variables instead of categorical (https://github.com/ribeirots/RILepi). This modification was preferred in light of the small counts of specific genotypes due to low heterozygosity and genomic ancestry skew present in our RIL panels.

The initial QTL mapping is based on the comparison of a model with a single QTL *versus* a model without any QTL, performed repeatedly one locus at a time. For each locus, we obtained a logarithm of the odds (LOD)

score as the ratio between the log-likelihoods of the model with one QTL over the model without a QTL. Because both RIL sets had average ancestral frequencies skewed against the Zambian ancestry (68% Ethiopia ancestry and 75% France ancestry), we excluded the windows with an ancestry bias of 90% or greater for either ancestry to avoid the effects of extremely uneven sample sizes on the model. The genome-wide significance of a QTL was tested using 10,000 permutations, in which trait values were randomly shuffled. From each permutation, we saved the highest LOD score obtained for any window (excluding windows with ancestry skew of 90% or greater) and empirical LOD scores from each window were compared to the null, permutation LOD score distribution. We considered the windows whose LOD score's p-value was lower than 0.10 as putative QTLs (*i.e.* a less than 10% chance of having a QTL signal of that strength anywhere in the genome). A p-value threshold of 0.10 was chosen aiming to include a sufficient number of QTLs to be tested for epistasis. Each putative QTL surrounded by windows with a lower LOD score was considered a putative QTL peak. If two putative QTL peaks were not separated by a window with a LOD score lower than the minor peak's LOD score minus 1.5, the minor peak was removed. Finally, the remaining putative QTL peaks were filtered to remove any peak within 10 cM of a QTL peak with a higher LOD score, in order to increase the independence of QTLs used in epistasis testing. The confidence interval for these QTLs

was defined as all contiguous windows with a LOD higher than the QTL's peak LOD score minus 1.5.

Candidate Genes and Population Genetics

To identify genes potentially underlying the adaptive traits, we scanned our QTL regions for signatures of local adaptation based on populations genetics. We used sequenced genomes from inbred lines from each of the studied populations to calculate a haplotype-based statistics: χ_{MD} (Lange & Pool 2016) and two F_{ST} statistics (using Reynolds *et al*. 1983), $F_{ST\_FullWin}$ (window-wide F_{ST} calculated using all the SNPs in the window) and F_{ST_MaxSNP} (the highest F_{ST} value from a single SNP within the window), that have power to distinct power to detect distinct kinds of selective events (da Silva Ribeiro *et al*. 2022). We defined as outlier windows those that fell within the top 1% of windows on the same chromosome arm for any of the above three statistics. We focused on regions with recombination rates generally above 0.5 cM/Mb (Comeron *et al*. 2012) due to more localized signatures of selection: 2.3–21.4 Mb of the chromosome X, 0.5–17.5 Mb of arm 2L, 5.2–20.8 Mb of arm 2R, 0.6–17.7 Mb of arm 3L, and 6.9–26.6 Mb of arm 3R. We selected as candidate genes all genes that overlapped with the outlier windows as well as the first gene up- and down-stream of the outlier window, to account for instances in which the target of selection is in a regulatory regions outside the gene region. We also identified other genes

within these QTLs with a known or potential association with the phenotype.

Epistasis Testing

Epistasis tests were performed between each focal QTL and all of the remaining genomic windows outside its confidence interval. The epistasis test for each window pair consisted of an interaction LOD score obtained as the ratio between the log-likelihoods of the model with two QTLs and their interaction (y ~ Q1 + Q2 + Q1:Q2) *versus* a model including both QTLs but no interaction (y ~ Q1 + Q2). The interaction LOD score for each pair was compared against permutations. Permutations were specific for each QTL analyzed. We shuffled the phenotype, fixed the focal QTL (Q1), and calculated the interaction LOD score of Q1 against all other windows in the genome. The highest genomewide interaction LOD score from each permutation was kept and the empirical interaction LOD scores were compared against this null, permuted interaction LOD distribution.

Epistasis Meta-analysis

We combined the results from all the epistasis tests across mapping crosses and phenotypes to test the hypothesis of whether adaptive QTLs have epistatic interactions. We collected the set of epistasis p-values from each QTL and its most likely interactor, which quantify the probability of obtaining any interaction term between that QTL and any partner locus in

the genome with an interaction LOD score as high as the one observed. We used Fisher's combined probability test to ask if the p-values from each QTL and their most likely interacting locus reject the null hypothesis that there is no epistasis. The p-values should follow a uniform distribution if the null hypothesis is true. Therefore, we performed an additional step to investigate how many of the lowest p-values would need to be removed from the data set to obtain a median value near 0.50.

Epistasis Power Analysis

We calculated the power of our approach and our data sets to detect an interaction by simulating phenotypes with different interaction strengths. We chose one empirical QTL with intermediate sample size and estimated the additive QTL effect among the detected QTL to serve as the basis for the simulations. The QTL chosen was detected for the abdominal background pigmentation at 25 °C, it had an estimated effect size equal to 0.035, mean trait value equal to 0.3485, and a standard deviation equal to 0.0697. The observed genotype distribution of the QTL peak window was used for the simulations. Our null simulations generated phenotypes based on a normal distribution with the observed trait mean and standard deviation, and the simulated values were modified based on the trait effect and individual genotypes (1x effect for heterozygous and 2x effect for non-Zambian homozygous). We then calculated the interaction LOD score of

the focal peak window against all other genome windows outside the original QTL's confidence interval and kept the highest genomewide interaction LOD score obtained. We repeated this process 10,000 times to obtain a null (no interaction) distribution of interaction LOD scores.

To simulate epistasis, we used the genotype distribution of the window most likely to be interacting with our focal QTL: the window with the highest empirical interaction LOD score. We modified the effect of the focal QTL on the simulated phenotype based on the genotype of the interacting window. If the genotype of the interacting window was Zambia homozygous, the original effect was not modified, if it was non-Zambian homozygous, the original effect was multiplied by the interacting factor I, and if the genotype of the interacting window was heterozygous, the original effect was intermediate, (i.e. multiplied by $(1+I)/2$). We simulated positive epistasis effects (original effect increased by the presence of the epistatic allele) ranging from 1.125 to 4. For diminishing returns negative epistasis, we simulated negative epistasis effects (original effect decreased by the presence of the epistatic allele) from 0.889 (1/1.125) to 0.5. For sign epistasis (the original effect changes in sign, e.g. instead of producing larger trait it produces smaller traits), we examined effects from -0.5 to -2. We also simulated masking epistasis (the original effect is nullified in the presence of the epistastic allele) with effect $I = 0$.

Each interaction simulation was compared against the null distribution and a p-value was calculated as the proportion of null simulations with an interaction LOD score higher than the simulated interacting LOD score. Power was calculated for each interaction effect *I* as the proportion of simulations that had p-value lower than 0.05.

We also calculated the power of our meta analysis based on 13 QTLs' interaction p-values. For each interaction effect, we sampled 13 interaction LOD score p-values and calculated (1) how often the fisher combined p-value was as extreme or greater than the one we observed and (2) how often at least one p-value was lower than the lowest p-value we obtained for our empirical data.

Reaction Norms

To investigate genotype-by-environment interactions for pigmentation traits, we combined phenotypic data from RILs raised at the two temperature treatments (15 °C and 25 °C) with the genotype of each non-overlapping QTL. To test whether the genotype-by-environment interaction was significant, we used a linear model with the phenotype as the response variable and the genotype, environment, and genotype-by-environment interaction as the explanatory variables.

Results

Genotype and phenotype data

Whole genome sequences for the RILs had mean depth of coverage of 3.3X (S.D. 1.96) and 7.58X (S.D. 9.09) per site for EF and FR RILs, respectively, which should be more than sufficient to call population ancestry in large chromosomal blocks. Ancestries for both RIL sets were skewed toward lower Zambian ancestry, with averages of 67.84% Ethiopian ancestry and 77.88% French ancestry. After five generations of inbreeding, a low level of parental strain heterozygosity was present in the final panel: 8.71% and 5.3% on the Ethiopian and French panels, respectively (at the level of genomic windows). The sample size varied for each trait based on the number of RILs successfully phenotyped (Table 1).

Correlations among the analyzed traits were examined for each RIL panel, which may indicate pleiotropy or else linkage of causative alleles. Among pigmentation traits, the thoracic mesopleuron trait was strongly correlated with abdominal background (Table S1). Stripe width showed moderate, significant correlations with both of those traits at 25 °C but weaker correlations at 15 °C. Abdominal background showed a non-significant correlation between the two temperatures, while the other two traits showed moderate, significant correlations between temperatures. In light of the correlations observed among many pigmentation traits, overlapping QTLs for pigmentation traits were not treated as independent

QTLs, and only the QTL with the highest LOD score was used in downstream analyses. No pigmentation trait was significantly correlated with wing length (Table S1), the only other trait scored among the Ethiopian RIL panel. Among traits scored from the French RIL panel, the song pattern was weakly correlated with ethanol resistance, $r^2 = 0.13$ (Table S2).

Single locus QTL mapping

We identified potential QTLs for eight out of the thirteen investigated traits (Figure 1; Table 1). The traits with no detected QTLs were A4 Background at 25C, all the three pigmentation plasticity traits, and cold tolerance. The number of QTLs per trait ranged up to five, for wing length, with other traits yielding one or three QTLs (Figure 1, Table 1). Twelve of the fourteen QTLs were significant at $p < 0.05$, while two marginally significant QTLs - one for A4 Background at 25 °C ($p = 0.0591$) and one QTL for ethanol resistance ($p = 0.0693$) – were also considered in downstream analyses (Table 2).

Of the five pigmentation QTLs, the two identified for stripe ratio (one at 25 °C and one at 15 °C) overlapped, and only the one with the highest LOD score (15 °C) was kept for meta-analysis (Table S3). Confidence intervals for the two QTLs detected for mesopleuron (one at 25

°C and one at 15 °C) were separated by just over 600 kb, so they were kept as two independent QTLs for the meta-analysis (Table 2).

The effect of all QTLs (Table 2) was in the expected direction: substituting alleles of the derived population resulted in phenotypic changes in the direction of the derived populations. This is congruent with a scenario where these QTLs are contributing to locally adaptive trait changes (involving either the measured traits or else pleiotropically correlated traits under selection), given that if the phenotypic difference were neutral, we would expect some QTLs to show an effect in the opposite direction (Orr 1998).

Outlier and Candidate Genes within QTLs

To generate hypotheses for possible causative genes underlying trait changes, we used population genetic summary statistics to identify genes within our QTLs that are potentially under local adaptation. Specifically, we flagged genes associated with windows that fell within the chromosome arm's top 1% of windows for window F_{ST}, maximum SNP F_{ST}, or the haplotype statistic χ_{MD} (Table 3). These local adaptation scans were not conducted in low recombination QTL regions due to the expected lack of gene-scale resolution of selection signals in such genomic intervals. Since a striking signal of local adaptation is not guaranteed at loci underlying these

strong population trait differences, we also noted the presence of other functionally relevant candidate genes within each QTL.

The mesopleuron pigmentation QTL detected at 25 °C included the gene *burs* (Dewey *et al*. 2004), implicated in cuticle tanning and hardening. The two stripe ratio QTLs, at the beginning of chromosome X, both included the candidate gene *Hr4*, a gene that when suppressed results in reduced pigmentation (Rogers *et al*. 2014). One also included the gene *yellow*, canonically implied in pigmentation variation and evolution (e.g. Massey & Wittkopp 2015). This pair of QTLs overlapped others previously detected for the same trait in distinct Ethiopia-Zambia crosses (Bastide *et al*. 2016). The gene *ebony*, at which a soft sweep signal associated with a pigmentation QTL was previously characterized for Ethiopian *D. melanogaster* (Bastide *et al*. 2016), fell just between the closely-located Meso15C and Meso25C peaks.

At least one gene related to wing development was found within each of the five wing length QTLs. Within the distal 2L QTL, our population genetic outliers included the *Drosophila* target of rapamycin gene (*Tor*), which plays a key role in insulin signaling and growth regulation, and has been tied to wing size and development specifically (Parker & Struhl 2015). Outliers inside this QTL also included the potential wing regulators *Vha68-2* and *Ube4B* (Krupp *et al*. 2005, Blanco *et al*. 2010, Okada *et al*. 2016, López-Varea *et al*. 2021). Within the low recombination distal 2L wing QTL,

among the few genes present was *Slmap*, a hippo signaling gene shown to alter wing size (Zheng *et al*. 2017). Within the 3L wing QTL, population genetic outliers included *klu*, which regulates cell size and proliferation, and is associated with wing size defects (Schertel *et al*. 2015). Within the 3R QTL, outliers included *pnt*, related to wing morphogenesis (Dworkin & Gibson, 2006, Bejarano *et al*. 2008, Paul *et al*. 2013). The X-linked wing QTL on chromosome X included an outlier gene possibly involved in wing development, *shakB* (Krishnan *et al*. 1993). Three of those QTLs (X, 2L proximal, and 3L) overlap with previously detected wing length QTLs, each from a different Ethiopia-Zambia cross out of four previously analyzed mapping crosses (Sprengelmeyer *et al*. 2020).

For the ethanol resistance QTLs, we identified two population genetic outlier genes (*htt* and *trv*) previously implicated in response to ethanol within the distal QTL on chromosome 3R (Fochler *et al*. 2017). Outlier genes detected within the other two QTLs had no annotated connection to ethanol. These QTLs did not overlap with those previously detected for the same trait from distinct France-Zambia crosses (Sprengelmeyer *et al*. 2021), in line with the genetically variable basis of this trait identified by that study.

Overall, we identified both novel genes underlying the studied traits and genes that had already been implicated with their respective trait. The identification of genes already known to underlie these traits, although not

novel, provides evidence that they may have been targeted by local adaptation, as well as offerings support for the approach we used here.

Epistasis

To investigate whether there is evidence of epistasis among adaptive loci, we performed an epistasis scan for each single QTL and a meta-analysis combining the results across non-correlated traits. We did not identify any individually significant epistatic QTL pairs. The lowest genome-wide interaction LOD score p-value we obtained was 0.183, for a wing length QTL (Table 2).

Our meta-analysis of the combined non-overlapping single QTLs also supports the hypothesis that there is no strong epistasis involving adaptive loci. By combining the interaction p-values for all non-overlapping QTLs (where each p-value denotes the probability of that QTL having any interaction LOD effect that strong in the genome, based on 10,0000 permutations), we obtained a p-value of 0.763) using Fisher's combined probability test. This result indicated that the observed p-values do not deviate from the null expectation of uniformly distributed p-values expected without epistasis involving adaptive loci).

Because our sample sizes of RILs may not be suitable for detecting all magnitudes of epistasis, we conducted a power analysis to indicate the strength of our negative results, and to indicate if there was a parameter

space of epistasis that we could confidently rule out based on these results. Based on our power analysis, our method had a high power (over 80%) to detect an epistatic QTL pair in scenarios of sign epistasis, in which the interaction effect changed the direction of the main effect, as well as high power to detect strong positive epistasis in which the interaction effected increased the main effect at least 2.5-fold (Table S4). We also tested the power of our meta-analysis by re-sampling 13 p-values from simulated data sets of different interaction effects. This power analysis indicated that we should have had power to detect similar scenarios of strong epistasis as indicated for the lowest p-value analysis above (*i.e.* sign epistasis and positive epistasis at least doubling the main effect, and that we would likely have detected masking epistasis as well, in which the main effect was erased by the modifier). Therefore, the lack of detected epistatic interactions from our empirical data supports the hypothesis that such strong forms of epistasis did not play an important role in these instances of evolution.

Phenotype plasticity for pigmentation QTLs

Fly pigmentation is known for its temperature-based plasticity; flies reared at lower temperatures develop darker phenotypes than those raised at warm temperatures (David *et al*. 1990). Our results recapitulated this behavior, in that average pigmentation traits were higher (*i.e.* darker) at 15 ˚C (Figure 3). We identified one overlapping QTL for stripe ratio at both

temperatures (Table 2). Based on the trait correlation and the shared QTL, at least part of the genetic architecture underlying adaptive melanism appears to be shared between temperatures.

The reaction norm of four pigmentation QTLs also showed that flies raised at 15 °C showed darker phenotypes for all traits and even darker phenotypes for the Ethiopia homozygotes, as expected (Figure 3). We also found significant genotype-by-environment interactions for two of these QTLs – A4 Background and stripe ratio (Figure 3C-D). The change in stripe ratio (Figure 3D) was greater for flies with the homozygous Ethiopia genotype. Here, the more derived-like environment appears to enhance the phenotypic consequences of the Ethiopian pigmentation variants. In contrast, for the A4 Background QTL, the phenotype means at 15 °C only show a small difference between homozygous genotypes, compared to the much larger difference at 25 °C, which is congruent with this QTL being found only at 25 °C.

Discussion

We report the creation of two new *Drosophila melanogaster* RIL panels established from crosses between single inbred lines from Ethiopia and Zambia, and from France and Zambia. We used these new data sets, together with an updated QTL mapping approach focusing on additive effects, to investigate the genetic architectures of multiple traits that have

evolved become differentiated between *D. melanogaster* populations. We did not find evidence for strong epistasis underlying their evolution, but genotype-by-environment interactions were supported for two of the three traits investigated.

Genetic architectures underlying trait evolution

Our QTL mapping results indicate genetic architectures for most traits involving detectable loci with non-trivial effect sizes (5.3% to 21.4%; Table 2), consistent with past bulk mapping studies for some of these traits (Bastide *et al*. 2016, Sprengelmeyer *et al*. 2020, Sprengelmeyer *et al*. 2021). In combination with those studies, our results reinforce previous findings that evolved populations maintain persistent variability in the genetic basis of pigmentation traits, wing length, and ethanol resistance. We did not find significant QTLs for cold tolerance or thermal plasticity of pigmentation, which could reflect either a more complex genetic basis of those traits or else greater non-genetic variance in those assays.

We have also identified genes within the detected QTLs that might be underlying the relevant traits, including genes previously known to be related to the phenotype and, with the use of population genetic signatures of selection, novel candidate genes (Table 3). The identified genes represent viable hypotheses for contributors to the genetic architecture of traits that appear to have evolved due to local adaptation targeting either

these or pleiotropically correlated traits. Additional studies using functional approaches such as genome editing will be needed to confirm the roles of these genes in the evolution of the respective traits.

Lack of evidence for epistasis

Despite the pervasiveness of epistasis underlying the genetic basis of complex traits (Huang *et al*. 2012, Mackay 2014), we did not find evidence of epistasis involving any of the detected adaptive loci (Table 2). In light of our power analysis (Figure 2), our empirical results primarily argue against the presence of modifier variants within our RIL panels that trigger sign epistasis, complete masking epistasis, or strong positive epistasis with regard to the main effect loci. In contrast, our study does not speak to the presence of quantitatively negative epistasis (*e.g.* diminishing returns epistasis) or more moderate positive synergistic epistasis.

Previous studies have also found mixed results when investigating epistatic interaction among adaptive loci (Malmberg & Mauricio 2005). In some cases, the varied outcome for adaptive traits might reflect the transient nature of epistatic interactions in a population, given that the epistatic effects of two loci on a trait depend on the allele frequency of the interacting alleles. Interestingly, studies mapping QTLs for fitness-related traits (more directly related to reproduction and survival) have a more

consistent result, often showing more epistatic QTLs than additive QTLs (Malmberg & Mauricio 2005).

One of the difficulties in extrapolating the results of experimental crosses to natural populations is that experimental crosses start with a limited amount of genetic variation (Mackay 2014). In our case, the crosses were made between one inbred line from each parental population, which will exclude some variants present in the source populations while elevating some naturally rare variants to moderate frequency. However, our crosses have the advantage of focusing on a simplified genetic architecture, which may improve power to detect interactions among the variants present, in addition to clear inference of the parental origin of alleles genome-wide.

We highlight that our study likely underestimates the epistatic interactions affecting *D. melanogaster* adaptive traits, since we only had power to detect relative strong interactions (Figure 2, Table S4). Based on the limited numbers of RILs available, we could only detect strong additive QTLs, and we needed to focus our epistasis scan on interactions with those few strong QTLs. Future research with a larger sample size (e.g. Torgeman & Zamir 2023) would be helpful to achieve more extensive insights into the scope of epistatic interactions affecting traits such as these.

Genetic Architectures at Two Temperatures

Although the three body pigmentation traits we studied were somewhat correlated (Table S1), no shared QTL was noted between them (Table 2). Instead, we found two instances of overlapping or neighboring QTLs for the same trait measured at different temperatures, with 25 °C and 15 °C being more similar to the mean temperatures in the Zambian and Ethiopian populations, respectively. These results are compatible with the same or nearby genes being responsible for pigmentation differences at both temperatures.

Drosophila pigmentation is known to show thermal plasticity, and Ethiopian flies would have been expected to become somewhat darken on this basis alone, and yet they have also adapted genetically to become darker (Bastide *et al*. 2014). It is unclear whether this pigmentation plasticity is adaptive, which in itself can facilitate adaptive evolution (Ghalambor *et al*. 2007). Unlike larger ectotherms, it has been estimated that the solar gain in temperature from having dark pigmentation is at most a fraction of a degree for an organism with the small size and high surface to volume ratio of *Drosophila* (Willmer & Unwin 1981, Hirai & Kimura 1997). Across African populations of *D. melanogaster*, levels of ultraviolet radiation have been found to be stronger predictors of pigmentation than temperature (Bastide *et al*. 2014).

Gene-by-environment loci have recently been shown to be enriched for signatures of positive selection (Lea *et al*. 2022), supporting the hypothesis that environment-specific loci play an important role in local adaptation. Here, we did find that two QTLs underlying pigmentation evolution showed significant genotype-by-environment interactions (Figure 3), suggesting that the rate of their response to selective events might have changed upon colonization of colder environments. However, we did not find consistent patterns of either gene-by-gene or gene-by-environment interactions that met the predictions of cryptic variation. For example, the absence of epistatic interactions means that the Ethiopian alleles did not have their effect completely masked or reversed by another gene on the ancestral background. And in the case of gene-environment interactions, there was an enhanced QTL effect in the derived Ethiopia-like cool environment in just one of two significant cases (*i.e.* the stripe QTL). While cryptic variation may have played a role in the evolution of some or all of these traits, our results suggest that it may not have been pervasive with respect to the strongest effect changes.

Conclusion

While the genetic architecture underlying adaptive evolution may vary somewhat depending upon the trait, there remains a key interest in assessing potentially general patterns. In combination with past studies,

our results suggest that most adaptive trait changes involve variants of non-trivial effect size, and that these variants often do not initially reach fixation. Epistasis is an important component underlying complex traits, including in flies (Huang *et al*. 2012), but we did not find evidence that strong epistasis was relevant for the adaptive evolution of the studied loci. In light of the existence of other examples in which epistasis among adaptive loci have been shown (Malmberg & Mauricio 2005), our study highlights that the answer to the classic debate on whether epistasis is important to natural selection might vary case by case. Lastly, we found evidence of gene-by-environment interactions underlying pigmentation, stressing that the variability of genetic architectures can also be environment-dependent, which is particularly relevant in light of rapidly changing global environments.

Data availability

Source codes used in the analyses are available on GitHub (https://github.com/ribeirots/RILepi). Genomic sequences for the Recombinant Inbred Lines will be available on SRA.

Acknowledgments

We would like to thank the Stern Lab for the use of their facilities for recording fly songs, particularly Ben Arthur for the *D. melanogaster* song

model provided. This research was funded by NIH grants R01 GM127480 and R35 GM13630, and by NSF grant DEB 1754745.

References

Abramoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, *11*(7), 36–42.

Adams, M., McBroome, J., Maurer, N., Pepper-Tunick, E., Saremi, N. F., Green, R. E., Vollmers, C., & Corbett-Detig, R. B. (2020). One fly–one genome: chromosome-scale genome assembly of a single outbred *Drosophila melanogaster*. *Nucleic Acids Research*, 48(13), e75-e75. https://doi.org/10.1093/nar/gkaa450.

Ang, R. M. L., Chen, S. A. A., Kern, A. F., Xie, Y., & Fraser, H. B. (2023). Widespread epistasis among beneficial genetic variants revealed by high-throughput genome editing. *Cell Genomics*, 3(4). https://doi.org/10.1016/j.xgen.2023.100260.

Arthur, B. J., Ding, Y., Sosale, M., Khalif, F., Kim, E., Waddell, P., Turaga, S. C., & Stern, D. L. (2021). *SongExplorer: A deep learning workflow for discovery and segmentation of animal acoustic communication signals* (p. 2021.03.26.437280). bioRxiv. https://doi.org/10.1101/2021.03.26.437280.

Arthur, B. J., Sunayama-Morita, T., Coen, P., Murthy, M., & Stern, D. L. (2013). Multi-channel acoustic recording and automated analysis of Drosophila courtship songs. *BMC Biology*, *11*(1), 11. https://doi.org/10.1186/1741-7007-11-11.

Bank, C. (2022). Epistasis and adaptation on fitness landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 53, 457-479. https://doi.org/10.1146/annurev-ecolsys-102320-112153.

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., & Pool, J. E. (2016). A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics*, *204*(3), 1307–1319. https://doi.org/10.1534/genetics.116.192492.

Bastide, H., Yassin, A., Johanning, E. J., & Pool, J. E. (2014). Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan Africa. *BMC Evolutionary Biology*, *14*(1), 179. https://doi.org/10.1186/s12862-014-0179-y.

Bejarano, F., Luque, C. M., Herranz, H., Sorrosal, G., Rafel, N., Pham, T. T., & Milán, M. (2008). A gain-of-function suppressor screen for genes involved in dorsal–ventral boundary formation in the *Drosophila* wing. *Genetics*, 178(1), 307-323.. http://dx.doi.org/10.1534/genetics.107.081869.

Blanco, E., Ruiz-Romero, M., Beltran, S., Bosch, M., Punset, A., Serras, F., & Corominas, M. (2010). Gene expression following induction of regeneration in *Drosophila* wing imaginal discs. Expression profile of regenerating wing discs. *BMC Developmental Biology*, 10(1), 1-14. http://dx.doi.org/10.1186/1471-213X-10-94.

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112.

Clemens, J., Coen, P., Roemschied, F. A., Pereira, T. D., Mazumder, D., Aldarondo, D. E., Pacheco, D. A., & Murthy, M. (2018). Discovery of a new song mode in *Drosophila* reveals hidden structure in the sensory and neural drivers of behavior. *Current Biology*, *28*(15), 2400-2412.e6. https://doi.org/10.1016/j.cub.2018.06.011.

Cohan, F. M., & Graf, J.-D. (1985). Latitudinal cline in *Drosophila melanogaster* for knockdown resistance to ethanol fumes and for rates of response to selection for further resistance. *Evolution*, *39*(2), 278–293. https://doi.org/10.2307/2408362.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genetics*, *8*(10), e1002905. https://doi.org/10.1371/journal.pgen.1002905.

Crow, J. F. (2010). On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1241–1244. https://doi.org/10.1098/rstb.2009.0275.

da Silva Ribeiro, T., Galván, J. A., & Pool, J. E. (2022). Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation. *Genome Biology and Evolution*, 14(10), evac143. https://doi.org/10.1093/gbe/evac143.

David, J. R., Capy, P., & Gauthier, J.-P. (1990). Abdominal pigmentation and growth temperature in *Drosophila melanogaster*: similarities and differences in the norms of reaction of successive segments. *Journal of Evolutionary Biology*, *3*(5–6), 429–445. https://doi.org/10.1046/j.1420-9101.1990.3050429.x.

Dewey, E. M., McNabb, S. L., Ewer, J., Kuo, G. R., Takanishi, C. L., Truman, J. W., & Honegger, H. W. (2004). Identification of the gene encoding bursicon, an insect neuropeptide responsible for cuticle sclerotization and wing spreading. *Current Biology*, 14(13), 1208-1213.. http://dx.doi.org/10.1016/j.cub.2004.06.051.

Doust, A. N., Lukens, L., Olsen, K. M., Mauro-Herrera, M., Meyer, A., & Rogers, K. (2014). Beyond the single gene: How epistasis and gene-by-environment effects influence crop domestication. *Proceedings of the National Academy of Sciences*, 111(17), 6178-6183. https://doi.org/10.1073/pnas.1308940110.

Duchêne, E., Butterlin, G., Dumas, V., & Merdinoglu, D. (2012). Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. *Theoretical and Applied Genetics*, 124(4), 623-635. https://doi.org/10.1007/s00122-011-1734-1.

Dworkin, I., & Gibson, G. (2006). Epidermal growth factor receptor and transforming growth factor-β signaling contributes to variation for wing shape in *Drosophila melanogaster*. *Genetics*, 173(3), 1417-1431. http://dx.doi.org/10.1534/genetics.105.053868.

Fenster, C. B., Galloway, L. F., & Chao, L. (1997). Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology & Evolution*, *12*(7), 282–286. https://doi.org/10.1016/S0169-5347(97)81027-0.

Fisher, R. A. (1930). *The genetical theory of natural selection* (pp. xiv, 272). Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Flynn, J. M., Rossouw, A., Cote-Hammarlof, P., Fragata, I., Mavor, D., Hollins III, C., Bank, C., & Bolon, D. N. (2020). Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife*, 9, e53810. https://doi.org/10.7554/eLife.53810.

Fochler, S., Morozova, T. V., Davis, M. R., Gearhart, A. W., Huang, W., Mackay, T. F., & Anholt, R. R. (2017). Genetics of alcohol consumption in Drosophila melanogaster. *Genes, Brain and Behavior*, 16(7), 675-685. https://doi.org/10.1111/gbb.12399.

Fry, J. D., Nuzhdin, S. V., Pasyukova, E. G., & Mackay, T. F. (1998). QTL mapping of genotype–environment interaction for fitness in *Drosophila melanogaster*. *Genetics Research*, 71(2), 133-141. https://doi.org/10.1017/S0016672398003176.

Fuhrmann, N., Prakash, C., & Kaiser, T. S. (2023). Polygenic adaptation from standing genetic variation allows rapid ecotype formation. *Elife*, 12, e82824. https://doi.org/10.7554/eLife.82824.

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x.

Gibson, G., & Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics 5*, 681–690. https://doi.org/10.1038/nrg1426.

Gong, L. I., & Bloom, J. D. (2014). Epistatically interacting substitutions are enriched during adaptive protein evolution. *PLoS Genetics*, 10(5). https://doi.org/10.1371/journal.pgen.1004328.

Griswold, C. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, 96, 445-453. https://doi.org/10.1038/sj.hdy.6800822.

Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution*, *67*(12), 3501–3511. https://doi.org/10.1111/evo.12214.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genetics*, *4*(2), e1000008. https://doi.org/10.1371/journal.pgen.1000008.

Hirai, Y., & Kimura, M. T. (1997). Incipient reproductive isolation between two morphs of *Drosophila elegans* (Diptera: Drosophilidae). *Biological Journal of the Linnean Society*, 61(4), 501-513. https://doi.org/10.1111/j.1095-8312.1997.tb01804.x.

Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R., & Mackay, T. F. (2020). Genotype by environment interaction for gene expression in *Drosophila melanogaster*. *Nature Communications*, 11(1), 5451. https://doi.org/10.1038/s41467-020-19131-y.

Huang, Y., Lack, J. B., Hoppel, G. T., & Pool, J. E. (2021). Parallel and population-specific gene regulatory evolution in cold-adapted fly populations. *Genetics*, 218(3). https://doi.org/10.1093/genetics/iyab077.

Huang, W., Richards, S., Carbone, M. A., Zhu, D., Anholt, R. R. H., Ayroles, J. F., Duncan, L., Jordan, K. W., Lawrence, F., Magwire, M. M., Warner, C. B., Blankenburg, K., Han, Y., Javaid, M., Jayaseelan, J., Jhangiani, S. N., Muzny, D., Ongeri, F., Perales, L., … Mackay, T. F. C. (2012). Epistasis dominates the genetic architecture of *Drosophila*

quantitative traits. *Proceedings of the National Academy of Sciences*, *109*(39), 15553–15559. https://doi.org/10.1073/pnas.1213423109.

Juenger, T. E., Sen, S., Stowe, K. A., & Simms, E. L. (2005). Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica*, 123, 87-105. https://doi.org/10.1007/s10709-003-2717-1.

Krishnan, S. N., Frei, E., Swain, G. P., & Wyman, R. J. (1993). Passover: a gene required for synaptic connectivity in the giant fiber system of Drosophila. *Cell*, 73(5), 967-977. https://doi.org/10.1016/0092-8674(93)90274-T.

Krupp, J. J., Yaich, L. E., Wessells, R. J., & Bodmer, R. (2005). Identification of genetic loci that interact with cut during *Drosophila* wing-margin development. *Genetics*, 170(4), 1775-1795. http://dx.doi.org/10.1534/genetics.105.043125.

Lachaise, D., Cariou, M.-L., David, J. R., Lemeunier, F., Tsacas, L., & Ashburner, M. (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. In M. K. Hecht, B. Wallace, & G. T. Prance (Eds.), *Evolutionary Biology* (pp. 159–225). Springer US. https://doi.org/10.1007/978-1-4613-0931-4_4.

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., Langley, C. H., & Pool, J. E. (2015). The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*, *199*(4), 1229–1241. https://doi.org/10.1534/genetics.115.174664.

Lack, J. B., Monette, M. J., Johanning, E. J., Sprengelmeyer, Q. D., & Pool, J. E. (2016). Decanalization of wing development accompanied the evolution of large wings in high-altitude *Drosophila*. *Proceedings of the National Academy of Sciences*, *113*(4), 1014–1019. https://doi.org/10.1073/pnas.1515964113.

Lange, J. D., & Pool, J. E. (2016). A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. *Molecular Ecology*, 25(13), 3081-3100. https://doi.org/10.1111/mec.13671.

Lea, A. J., Peng, J., & Ayroles, J. F. (2022). Diverse environmental perturbations reveal the evolution and context-dependency of genetic effects on gene expression levels. *Genome Research*, 32(10), 1826-1839.https://www.genome.org/cgi/doi/10.1101/gr.276430.121.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997.

Lollar, M. J., da Silva Ribeiro, T., Stern. D., & Pool, J. E. (*in Prep*). Quantitative trait mapping of courtship song variation between male *Drosophila melanogaster* from two allopatric populations.

López-Varea, A., Vega-Cuesta, P., Ruiz-Gómez, A., Ostalé, C. M., Molnar, C., Hevia, C. F., Martín, M., Organista, M. F., de Celis, J., Culí, J., Esteban, N., & de Celis, J. F. (2021). Genome-wide phenotypic RNAi screen in the *Drosophila* wing: Phenotypic description of functional classes. *G3*, 11(12), jkab349. https://doi.org/10.1093/g3journal/jkab349.

Lowry, D. B., Lovell, J. T., Zhang, L., Bonnette, J., Fay, P. A., Mitchell, R. B., Lloyd-Reilley, J., Boe, A. R., Wu, Y., Rouquette Jr, F. M., Wynia, R. L., Weng, X., Behrman, K. D., Healey, A., Barry, K., Lipzen, A., Bauer, D., Sharma, A., Jenkins, J., Schmutz, J., Fritschi, F. B., & Juenger, T. E. (2019). QTL × environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proceedings of the National Academy of Sciences*, 116(26), 12933-12941. https://doi.org/10.1073/pnas.1821543116.

Lunzer, M., Golding, G. B., & Dean, A. M. (2010). Pervasive cryptic epistasis in molecular evolution. *PLoS Genetics*, 6(10). https://doi.org/10.1371/journal.pgen.1001162.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, 35(1), 303-339. https://doi.org/10.1146/annurev.genet.35.102401.090633.

Mackay, T. F. C. (2014). Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, *15*(1), 22–33. https://doi.org/10.1038/nrg3627.

Malmberg, R. L., & Mauricio, R. (2005). QTL-based evidence for the role of epistasis in evolution. *Genetics Research*, *86*(2), 89–95. https://doi.org/10.1017/S0016672305007780.

Martin, G., & Lenormand, T. (2006). The fitness effect of mutations across environments: A survey in light of fitness landscape models. *Evolution*, 60(12), 2413-2427. https://doi.org/10.1111/j.0014-3820.2006.tb01878.x.

Massey, J. H., & Wittkopp, P. J. (2016). The genetic basis of pigmentation differences within and between *Drosophila* species. *Current Topics in developmental Biology*, 119, 27-61. https://doi.org/10.1016/bs.ctdb.2016.03.004.

Mathews, K. L., Malosetti, M., Chapman, S., McIntyre, L., Reynolds, M., Shorter, R., & Van Eeuwijk, F. (2008). Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theoretical and Applied Genetics*, 117, 1077-1091. https://doi.org/10.1007/s00122-008-0846-8.

McGuigan, K., & Sgro, C. M. (2009). Evolutionary consequences of cryptic genetic variation. *Trends in Ecology & Evolution*, 24(6), 305-311. https://doi.org/10.1016/j.tree.2009.02.001.

Okada, H., Ebhardt, H. A., Vonesch, S. C., Aebersold, R., & Hafen, E. (2016). Proteome-wide association studies identify biochemical modules associated with a wing-size phenotype in *Drosophila melanogaster*. *Nature Communications*, 7(1), 12649. http://dx.doi.org/10.1038/ncomms12649.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, *52*(4), 935–949. https://doi.org/10.1111/j.1558-5646.1998.tb01823.x.

Papa, R., Kapan, D. D., Counterman, B. A., Maldonado, K., Lindstrom, D. P., Reed, R. D., Nijhout, H. F., Hrbek, T., & McMillan, W. O. (2013). Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in *Heliconius erato*. *PLoS One*, 8(3), e57033. https://doi.org/10.1371/journal.pone.0057033.

Parker, J., & Struhl, G. (2015). Scaling the Drosophila wing: TOR-dependent target gene access by the Hippo pathway transducer Yorkie. *PLoS Biology*, 13(10), e1002274.. http://dx.doi.org/10.1371/journal.pbio.1002274.

Paul, L., Wang, S. H., Manivannan, S. N., Bonanno, L., Lewis, S., Austin, C. L., & Simcox, A. (2013). Dpp-induced Egfr signaling triggers postembryonic wing development in *Drosophila*. *Proceedings of the National Academy of Sciences*, 110(13), 5058-5063. http://dx.doi.org/10.1073/pnas.1217538110.

Pitchers, W., Pool, J. E., & Dworkin, I. (2013). Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution*, *67*(2), 438–452. https://doi.org/10.1111/j.1558-5646.2012.01774.x.

Pool, J. E. (2015). The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Molecular Biology and Evolution*, *32*(12), 3236–3251. https://doi.org/10.1093/molbev/msv194.

Pool, J. E., & Aquadro, C. F. (2006). History and structure of sub-Saharan populations of *Drosophila melanogaster*. *Genetics*, *174*(2), 915–929. https://doi.org/10.1534/genetics.106.058693.

Pool, J. E., Braun, D. T., & Lack, J. B. (2017). Parallel evolution of cold tolerance within *Drosophila melanogaster*. *Molecular Biology and Evolution*, *34*(2), 349–360. https://doi.org/10.1093/molbev/msw232.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4), R208-R215. https://doi.org/10.1016/j.cub.2009.11.055.

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105(3), 767-779. https://doi.org/10.1093/genetics/105.3.767.

Rogers, W. A., Grover, S., Stringer, S. J., Parks, J., Rebeiz, M., & Williams, T. M. (2014). A survey of the trans-regulatory landscape for *Drosophila melanogaster* abdominal pigmentation. *Developmental Biology*, 385(2), 417-432. https://doi.org/10.1016/j.ydbio.2013.11.013.

Rutherford, S. L., & Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, 396(6709), 336-342. https://doi.org/10.1038/24550.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics, 14, 807-820. https://doi.org/10.1038/nrg3522.

Schertel, C., Albarca, M., Rockel-Bauer, C., Kelley, N. W., Bischof, J., Hens, K., van Nimwegen, E., Basier, K., & Deplancke, B. (2015). A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating *Drosophila* wing development. *Genome Research*, 25(4), 514-523. https://doi.org/10.1101/gr.181305.114.

Schlötterer, C. (2023). How predictable is adaptation from standing genetic variation? Experimental evolution in *Drosophila* highlights the central role of redundancy and linkage disequilibrium. *Philosophical Transactions of the Royal Society B*, 378(1877), 20220046. https://doi.org/10.1098/rstb.2022.0046.

Sprengelmeyer, Q. D., Mansourian, S., Lange, J. D., Matute, D. R., Cooper, B. S., Jirle, E. V., Stensmyr, M. C., & Pool, J. E. (2020). Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Molecular Biology and Evolution*, *37*(3), 627–638. https://doi.org/10.1093/molbev/msz271.

Sprengelmeyer, Q. D., & Pool, J. E. (2021). Ethanol resistance in *Drosophila melanogaster* has increased in parallel cold-adapted populations and shows a variable genetic architecture within and between populations. *Ecology and Evolution*, *11*(21), 15364–15376. https://doi.org/10.1002/ece3.8228.

Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*(7), 1204-1218. https://doi.org/10.1002/pro.2897.

Steiner, C. C., Weber, J. N., & Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, *5*(9), e219. https://doi.org/10.1371/journal.pbio.0050219.

Stitzer, M. C., & Ross-Ibarra, J. (2018). Maize domestication and gene interaction. *New Phytologist*, 220(2), 395-408. https://doi.org/10.1111/nph.15350.

Torgeman, S., & Zamir, D. (2023). Epistatic QTLs for yield heterosis in tomato. *Proceedings of the National Academy of Sciences*, *120*(14), e2205787119. https://doi.org/10.1073/pnas.2205787119.

Trindade, S., Sousa, A., Xavier, K. B., Dionisio, F., Ferreira, M. G., & Gordo, I. (2009). Positive epistasis drives the acquisition of multidrug resistance. *PLOS Genetics*, *5*(7), e1000578. https://doi.org/10.1371/journal.pgen.1000578.

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Vijendravarma, R. K., & Kawecki, T. J. (2013). Epistasis and maternal effects in experimental adaptation to chronic nutritional stress in *Drosophila*. *Journal of Evolutionary Biology*, 26(12), 2566-2580. https://doi.org/10.1111/jeb.12248.

Wade, M. J. (1992). Sewall Wright: Gene interaction and the shifting balance theory. In *Oxford surveys in evolutionary biology* (Vol. 8, pp. 35–35).

Whitlock, M. C., Phillips, P. C., Moore, F. B.-G., & Tonsor, S. J. (1995). Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, *26*(1), 601–629. https://doi.org/10.1146/annurev.es.26.110195.003125.

Willmer, P. G., & Unwin, D. M. (1981). Field analyses of insect heat budgets: reflectance, size and heating rates. *Oecologia*, 50, 250-255. https://doi.org/10.1007/BF00348047.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093%2Fgenetics%2F16.2.97.

Yeaman, S. & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, 65(7), 2123–2129. https://doi.org/10.1111/j.1558-5646.2011.01277.x.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x.

Zeng, Z. B., Kao, C. H., & Basten, C. J. (1999). Estimating the genetic architecture of quantitative traits. *Genetics Research*, 74(3), 279-289. https://doi.org/10.1017/S0016672399004255.

Zheng, Y., Liu, B., Wang, L., Lei, H., Prieto, K. D. P., & Pan, D. (2017). Homeostatic control of Hpo/MST kinase activity through autophosphorylation-dependent recruitment of the STRIPAK PP2A phosphatase complex. *Cell Reports*, 21(12), 3612-3623. https://doi.org/10.1016/j.celrep.2017.11.076.

Table 1. Summary of mapped traits. Listing the mapping cross used for each trait (EF = Ethiopia, FR = France), the number of phenotyped lines for each trait, the number of QTLs, and the number of QTLs with epistatic interaction for each phenotypic trait.

RIL panel	Trait	Sample size	# QTLs
EF	A4 Background 25 °C	278	1
EF	Stripe Ratio 25 °C	278	1
EF	Mesopleuron 25 °C	278	1
EF	A4 Background 15 °C	190	0
EF	Stripe Ratio 15 °C	190	1
EF	Mesopleuron 15 °C	190	1
EF	A4 Background (Plast.)	178	0
EF	Stripe Ratio (Plast.)	178	0
EF	Mesopleuron (Plast.)	178	0
EF	Wing Length	268	5
FR	Cold Tolerance	304	0
FR	Ethanol Resistance	277	3
FR	Song Pattern	294	1

Table 2. Characteristics of detected QTLs for all analyzed phenotypes. The single QTL coefficient shows the magnitude of phenotypic change by substituting one allele of the derived population ancestry at that locus, which is shown on the scale of the measured phenotype. The correlation coefficient r^2 shows how much of the phenotypic variation Locus 1 explains (how well the data fits the model). QTL p-value and epistasis p-value were obtained based on 10,000 permutations. * an estimated p-value of zero

means the empirical value was more extreme than any permutation. C.I. is the confidence interval for Locus 1. Further details of each QTL are given in Table S3.

RIL panel	Trait	LOD score	Trait Coeff.	r^2	QTL p-value	C. I.	Epist. p-value
EF	A4Back 25C	3.274	0.012	0.053	0.0591	Chr3R:21,614,478..22,512,754	0.736
EF	Mesopl 15C	5.819	0.012	0.132	0.0006	Chr3R:16,750,589..16,945,416	0.699
EF	Mesopl 25C	5.890	0.014	0.093	0.0003	Chr3R:17,549,430..17,666,763	0.714
EF	Stripe 25C	6.335	0.035	0.100	0.0002	ChrX:0..2,141,596	0.643
EF	**Stripe 15C**	**9.924**	**0.071**	**0.214**	**0***	**ChrX:1,057,169.. 2,420,139**	**0.244**
EF	Wing	9.581	0.026	0.153	0*	Chr2L:12,902,781..13,046,286	0.322
EF	Wing	8.628	0.024	0.139	0*	Chr2L:22,740,849..23,011,543	0.672
EF	Wing	3.867	0.022	0.065	0.0183	Chr3L:10,353,468..11,407,946	0.444
EF	Wing	3.526	0.019	0.059	0.0369	Chr3R:18,722,896..19,232,530	0.941
EF	Wing	4.286	0.016	0.072	0.0073	ChrX:20,419,847.. 20,893,609	0.183
FR	EtOH	3.565	22.516	0.059	0.0368	Chr3L:10,829,907..11,166,414	0.208
FR	EtOH	3.253	27.506	0.054	0.0693	Chr3R:3,268,049..3,791,969	0.853
FR	EtOH	3.740	19.166	0.061	0.0246	Chr3R:23,770,233..24,580,087	0.476
FR	Song	9.111	0.019	0.133	0*	Chr3R:21,997,113..22,171,358	0.273

Table 3. QTL traits, confidence interval, and candidate genes detected by a priori functional knowledge or as population genetics outliers (n/a indicates that QTL was in a region of low recombination and was not analyzed for signatures of selection).

Trait	QTL Confidence Interval	Candidate Genes	Genes with Signatures of Selection
A4Back 25C	Chr3R:21,614,478.. 225,12,754		*TwdlQ, beat-vii, scrib*
Meso 15C	Chr3R:16,750,589.. 16,945,416		
Meso 25C	Chr3R:17,549,430.. 17,666,763	*burs*	
Strp 25C	ChrX:0..2,141,596	*Hr4, y*	n/a
Strp 15C	ChrX:1,057,169.. 2,420,139	*Hr4*	n/a
Wing	Chr2L:12,902,781.. 13,046,286		*A16, Vha68-2, ACXE, CG16800, Tor, Ube4B*
Wing	Chr2L:22,740,849.. 23,011,543	*Slmap*	n/a
Wing	Chr3L:10,353,468.. 11,407,946	*Cpr67Fa2, FoxK, Rbfox1, scyl, wls*	*Fad2, CG32079, klu*
Wing	Chr3R:18,722,896.. 19,232,530	*Cow, Gbp3, hh, lrk1, p53, pnt, Rassf*	*Pnt*
Wing	ChrX:20,419,847.. 20,893,609		*CG15450, Cyp6v1, hydra, shakB, CG1835, CG11227*
EtOH	Chr3L:10,829,907.. 11,166,414	*Aps*	*CG42831, tna*
EtOH	Chr3R:3,268,049.. 3,791,969		n/a
EtOH	Chr3R:23,770,233..24,580,087	*fkh*	*trv, htt, CG34354, CG43125, snu*
Song	Chr3R:21,997,113.. 22,171,358		

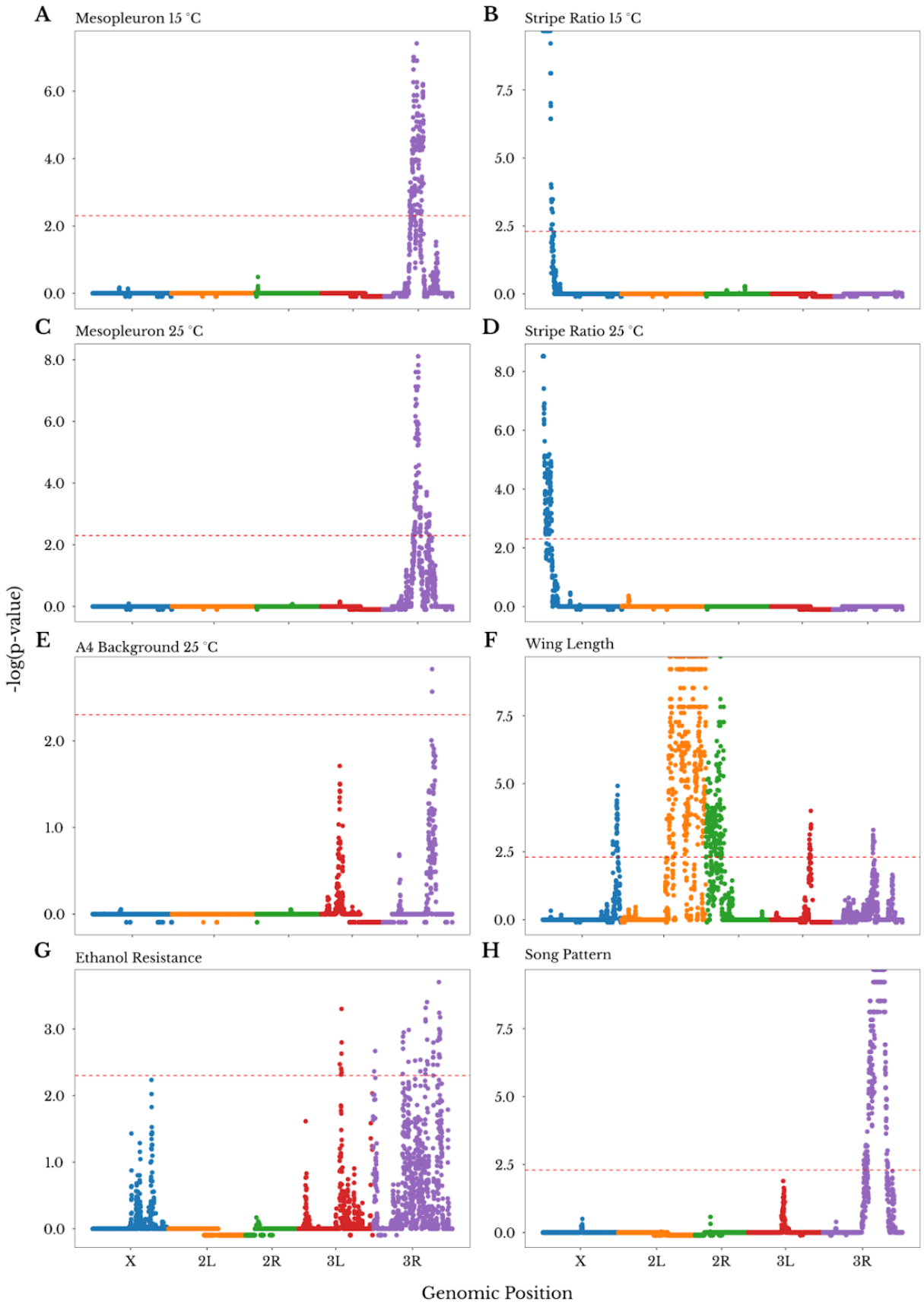

A
Mesopleuron 15 °C
B
Stripe Ratio 15 °C
C
Mesopleuron 25 °C
D
Stripe Ratio 25 °C
E
A4 Background 25 °C
F
Wing Length
G
Ethanol Resistance
H
Song Pattern
-log(p-value)
X
2L
2R
3L
3R
Genomic Position

Figure 1. The number of QTLs for each trait ranged from 0 to 5. In this figure, only the traits with at least one identified QTL are shown; the traits without QTLs (Table 1) can be seen in Figure S1. Each panel shows the -log of the p-value for the LOD score of the genomic windows. A4 Background (Figure 1E) is shown here with no QTL above the threshold, but one marginally significant QTL (p-value = 0.053, Table 2). Windows filtered out for ancestry skew are given a value of -0.25. The red dashed line represents the 0.05 p-value cutoff based on 10,000 permutations. The color of the dots represents the chromosome arm of each genomic window. Note that for wing length, a single QTL spans a broad low recombination centromeric region between 2L and 2R.

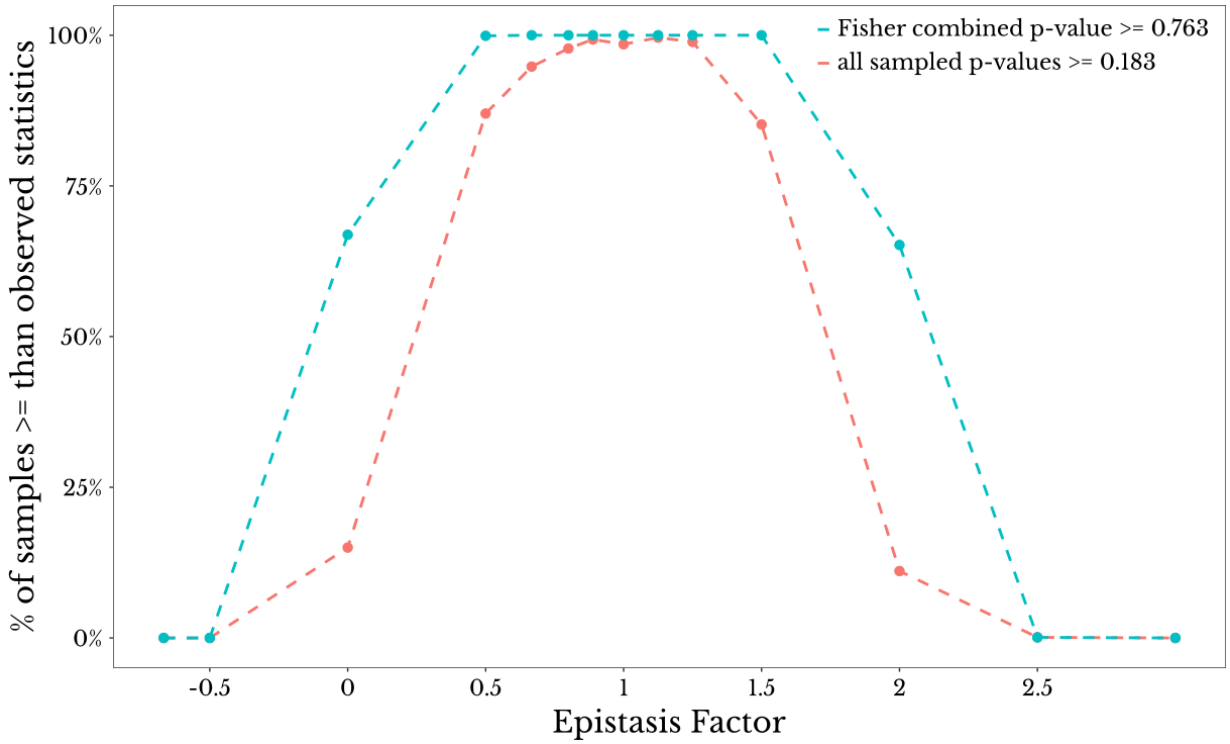


Figure 2. Power analysis indicating that strong positive epistasis or sign epistasis should have been detectable for the empirical data. Power analyses based on randomly permuted individual genotypes and simulated epistatic effects (see Materials and Methods) were conducted to assess the power to detect varying models of epistasis based on either the lowest empirical epistasis p-value (red series) or the Fisher-combined epistasis p-value across 13 analyzed QTLs (blue series). Here, the epistasis factor (*I*, x-axis) represents the multiplier that a modifier locus exerts on the primary QTL's effect. Thus, negative values represent sign epistasis, zero represents masking epistasis, values less than 1 indicate negative epistasis, and values above 1 indicate positive epistasis. The y-axis shows the proportion of occurrences out of the 1,000 resampled instances in which a given model of epistasis yielded a lowest p-value greater than or equal to the observed 0.183 (red), or else how often the combined Fisher p-value was greater than or equal to the observed 0.763 (blue). Results indicated that models of strong epistasis such as sign epistasis or strong positive epistasis should have been detectable from our data.

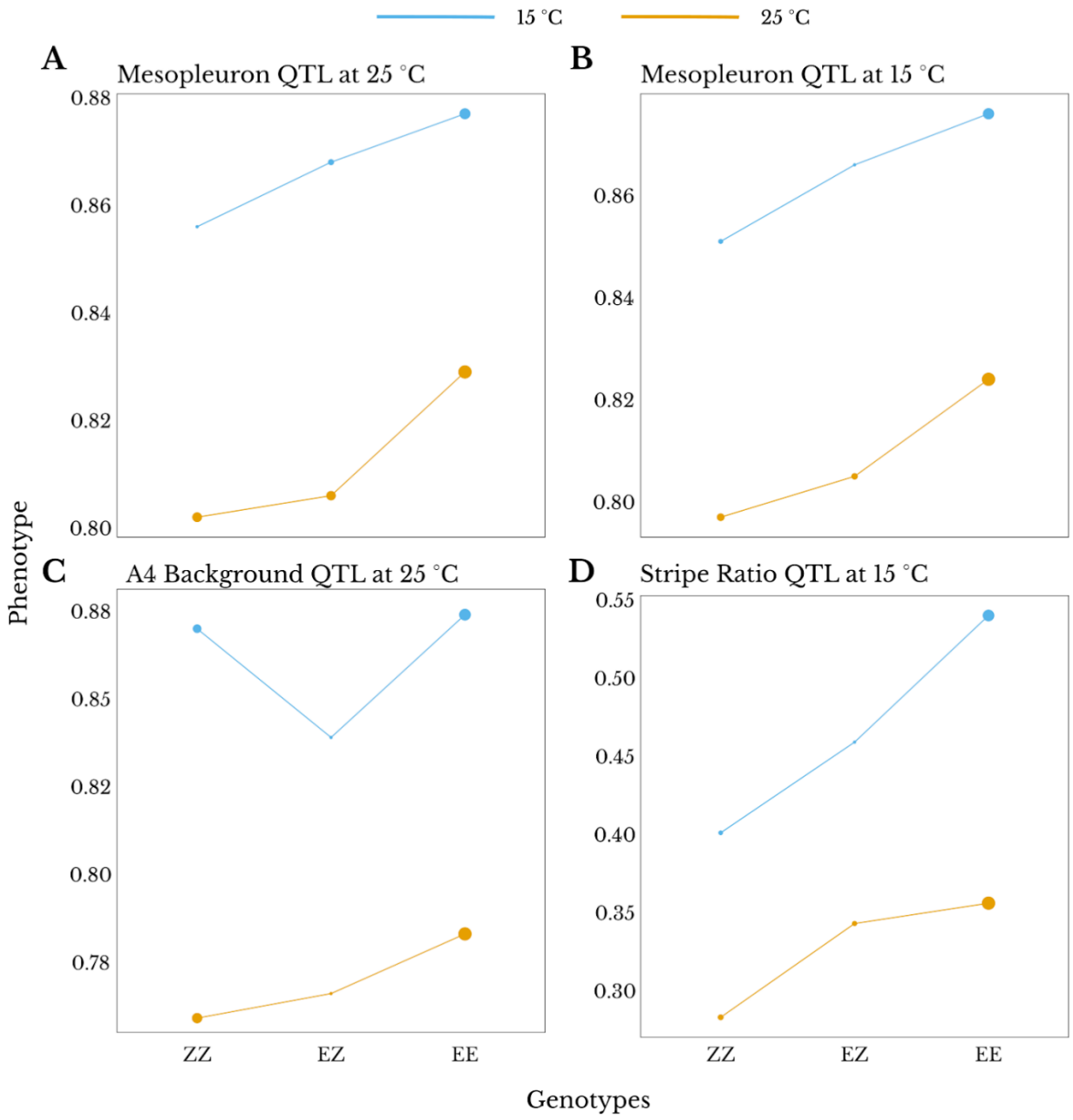


Figure 3. The reaction norms of the four non-overlapping pigmentation QTL show darker phenotypes for flies raised at 15 °C than at 25 °C and for flies with Ethiopia ancestry alleles, corroborating expectations. (A) Mesopleuron at 25 °C, QTL on chromosome 3 with peak position from 17,566,218 to 17,600,070. (B) Mesopleuron at 15 °C, QTL on chromosome

3R with peak position from 16,773,298 to 16,788,420. (C) A4 Background at 25C, QTL on chromosome 3R with peak from 21,614,478 to 21,628,049. (D) Stripe ratio at 15 ˚C, QTL on chromosome X with peak from 1,246,761 to 1,303,582. The y-axis shows the mean phenotype for each genotype and temperature treatment; a higher number is a darker phenotype. ZZ = Zambia homozygous, EZ = heterozygous, EE = Ethiopia homozygous.

Supplementary tables and figures

Table S1. Correlation among Ethiopia x Zambia RIL phenotypic traits. Wing = wing length. A4Back = Abdominal segment 4 background pimentation, Meso = Mesopleura pigmentation, Strp = Abdominal segment 4 stripe proportion. 15 = Trait measured at 15 ˚C, 25 = Trait measured at 25 ˚C, Pl = Plasticity (25 ˚C - 15 ˚C). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Back15	Meso15	Strp15	Back25	Meso25	Strp25	BackPl	MesoPl	StrpPl	Wing
Back15	1.000									
Meso15	0.687***	1.000								
Strp15	0.169*	-0.085	1.000							
Back25	0.133	0.155*	0.186*	1.000						
Meso25	0.197**	0.318***	0.130	0.800***	1.000					
Strp25	0.260***	0.209**	0.294***	0.460***	0.253***	1.000				
BackPl	-0.430***	-0.231**	0.075	0.838***	0.637***	0.262	1.000			
MesoPl	-0.265***	-0.366***	0.191**	0.698***	0.766***	0.140	0.782	1.000		
StrpPl	0.033	0.241***	-0.708***	0.156*	0.092	0.467	0.124	-0.073	1.000	
Wing	-0.085	-0.043	-0.043	0.033	0.047	-0.019	0.044	0.051	0.067	1.000

Table S2. Correlation among France x Zambia RIL phenotypic traits. Song = Pulse song fast proportion, Cold = Cold tolerance, EtOH = Ethanol resistance. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Song	Cold	Etoh
Song	1.000		
Cold	0.075	1.000	
EtOH	0.128*	-0.063	1.000

Table S3. List of QTL locations, statistical strength, and epistasis test results. RIl panel indicates the mapping cross used for each trait (EF = Ethiopia x Zambia, FR = France x Zambia). Phenotype column lists the trait measured, Wing = wing length. A4Back = Abdominal segment 4 background pimentation, Meso = Mesopleura pigmentation, Strp = Abdominal segment 4 stripe proportion. 15 = Trait measured at 15 °C, 25 = Trait measured at 25 °C, Pl = Plasticity (25 °C - 15 °C), Song = Pulse song fast proportion, Cold = Cold tolerance, EtOH = Ethanol resistance. The single QTL coefficient shows the magnitude of phenotypic change by substituting one allele of the derived population ancestry at that locus, which is shown on the scale of the measured phenotype. The correlation coefficient r2 shows how much of the phenotypic variation Locus 1 explains (how well the data fits the model). QTL p-value and epistasis p-value were obtained based on 10,000 permutations. Locus 2 indicates the locus in the pairwise comparison that yielded the lowest epistasis p-value, shown in in the epistasis p-value column. 0*: p-value < 0.0001.

RIL panel	Phenotype	LOD	Coeff	r2	p-value	Locus 1	Confidence Interval	Locus 2	Epistasis p-value
EF	A4back25	3.274	0.012	0.053	0.0591	Chr3R:21614478..21628049	Chr3R:21614478..22512754	2R:17045112..17062713	0.736
EF	Meso15	5.819	0.012	0.132	0.0006	Chr3R:16773298..16788420	Chr3R:16750589..16945416	X:19262570..19275538	0.699
EF	Meso25	5.890	0.014	0.093	0.0003	Chr3R:17566218..17600070	Chr3R:17549430..17666763	2L:11650732..11662916	0.714
EF	Strp15	9.924	0.071	0.214	0*	ChrX:1246761..1303582	ChrX:1057169..2420139	2L:2159941..2177260	0.244
EF	Strp25	6.335	0.035	0.100	0.0002	ChrX:795941..816187	ChrX:0..2141596	2L:8906629..8921342	0.643
EF	Wing	9.581	0.026	0.153	0*	Chr2L:12968914..12981303	Chr2L:12902781..13046286	2L:16273440..16293792	0.322
EF	Wing	8.628	0.024	0.139	0*	Chr2L:22740849..23011543	Chr2L:22740849..23011543	X:3996657..4059520	0.672
EF	Wing	3.867	0.022	0.065	0.0183	Chr3L:10803688..10816457	Chr3L:10353468..11407946	X:3289225..3301143	0.444
EF	Wing	3.526	0.019	0.059	0.0369	Chr3R:18903335..18914948	Chr3R:18722896..19232530	3R:18609961..18625178	0.941
EF	Wing	4.286	0.016	0.072	0.0073	ChrX:20773546..20787954	ChrX:20419847..20893609	3L:8741958..8757230	0.183
FR	EtOH	3.565	22.516	0.059	0.0368	Chr3L:11046891..11060714	Chr3L:10829907..11166414	2R:20261218..20316093	0.208
FR	EtOH	3.740	19.166	0.061	0.0246	Chr3R:23822228..23838090	Chr3R:23770233..24580087	3L:1888828..1903274	0.476
FR	EtOH	3.253	27.506	0.054	0.0693	Chr3R:3268049..3305900	Chr3R:3268049..3791969	2R:20261218..20316093	0.853
FR	Song	9.111	0.019	0.133	0*	Chr3R:22112128..22126990	Chr3R:21997113..22171358	3R:12326080..12341721	0.273

Table S4. Power to detected an interaction QTL based on the strength of the interaction factor. Power calculated as percentage of simulations with interaction LOD score p-value lower than the 5% null distributions of interaction LOD scores obtained in simulations without epistasis.

Epistasis Factor	Power
-2.000	1.00
-1.500	0.98
-1.250	0.93
-1.125	0.88
-1.000	0.81
-0.889	0.73
-0.800	0.66
-0.667	0.54
-0.500	0.38
0.000	0.07
0.500	0.00
0.667	0.00
0.800	0.00
0.889	0.00
1.000	0.00
1.125	0.00
1.250	0.00
1.500	0.00
2.000	0.07
2.500	0.37
3.000	0.80
3.500	0.98
4.000	1.00

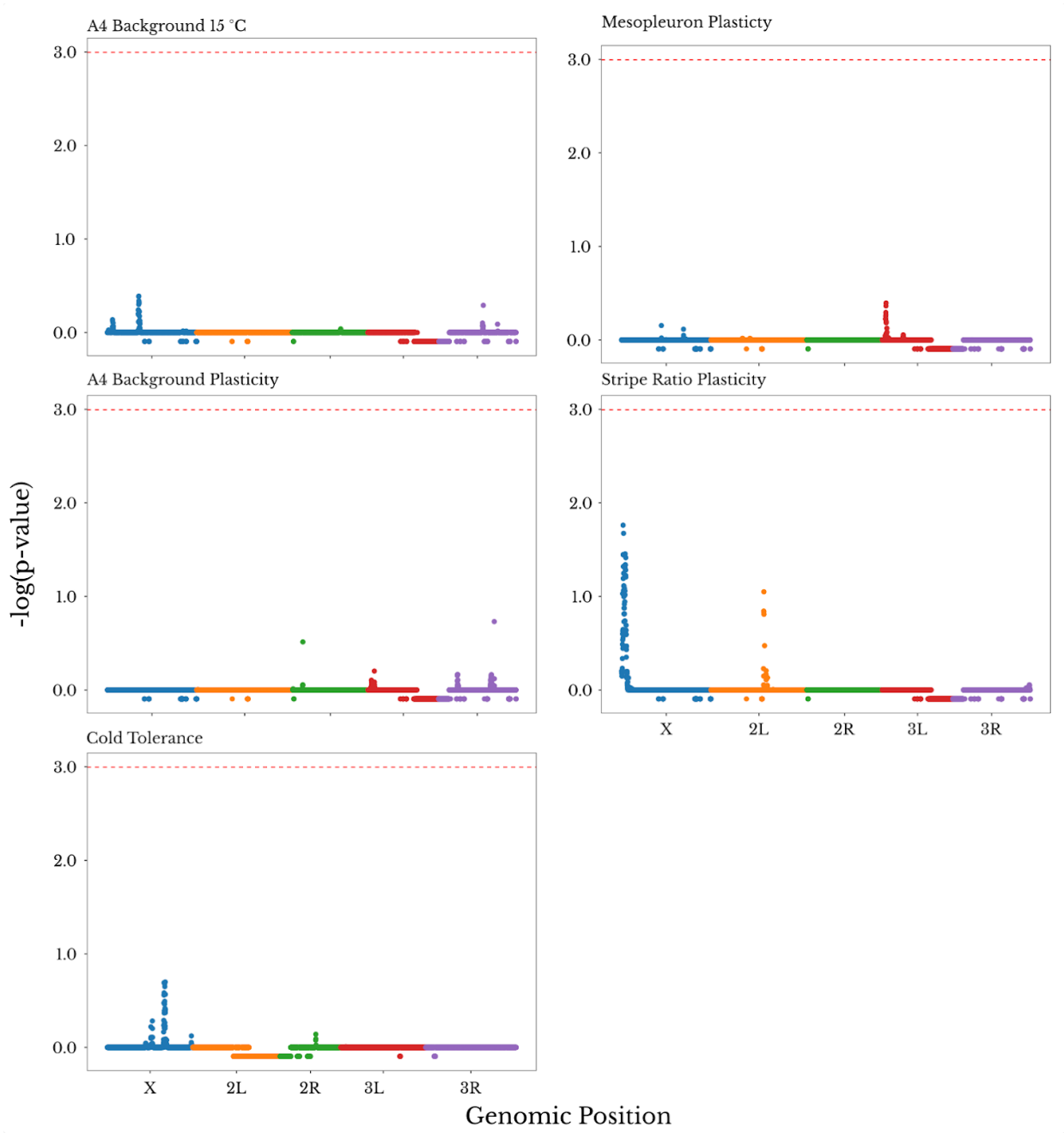


Figure S1. Genomewide LOD scores for traits without QTLs. Each panel shows the -log of the p-value for the LOD score of the genomic windows. Windows filtered out due to strong ancestry skew appear below zero. The red dashed line represents the 0.05 p-value cutoff based on 10,000

permutations. The color of the dots represents the chromosome arm of each genomic window.

Chapter 4: Adaptive variants underlying melanism in high altitude *Drosophila melanogaster* are polymorphic in both ancestral and derived populations

Co-author contribution acknowledgment

We would like to thank Jamie C. Freeman for performing the molecular biology work necessary to generate the genomewide data used in this project, as well as for assisting in the writing of the present text.

Abstract

Understanding the genetic basis of adaptation is a central question in evolutionary biology. Empirical studies have shown distinct adaptive genetic architectures among species, traits, and populations. Here, we perform extensive quantitative trait locus (QTL) mapping experiments focused on the darkest known population of *Drosophila melanogaster*, from high-altitude Ethiopia, to investigate the genetic architectures underlying this instance of melanic evolution. We mapped three distinct pigmentation traits in 21 mapping crosses between dark strains from Ethiopia and light strains from a Zambian population from the species' ancestral range. QTLs overlapping the canonical pigmentation genes *ebony*, *tan*, and *yellow* were each present in just under half of all mapping experiments, and tended to have stronger phenotypic effects. Some

additional QTLs overlapped with documented pigmentation genes, while other QTLs point to presently unknown contributors. We also performed mapping for a subset of crosses at a cooler, more Ethiopian-like temperature, which indicated thermally plastic effects on a minority of the QTLs that may have enhanced or resisted the evolution of melanism in Ethiopia. On average, we found that the Ethiopian and the Zambian parental strains involved in a cross were equally powerful determinants of the QTLs detected. These results are congruent with selection on relatively common pigmentation variants that were already present in the ancestral range, and rose moderately in frequency under local adaptation in Ethiopia but did not approach fixation. Thus, even for fly pigmentation traits often thought to have relatively simple molecular underpinnings, we find evidence that an abundance of standing genetic variation gave rise to persistently variable genetic architectures underlying adaptive traits in the evolved population.

Introduction

Understanding the genetic basis of adaptation has been a central question in evolutionary biology since the onset of the modern synthesis. The relative contribution of many loci of small effect *versus* a few loci of large effect to adaptive evolution is still a topic of debate (Fisher 1930, Pritchard *et al*. 2010, Rockman 2012, Schluter *et al*. 2021). Other factors can

also affect the genetic architecture of adaptation, such as the source of genetic variation and gene flow. The prevalence of new mutations *versus* standing variation underlying adaptation is still a central question in the field. Population genetics theory has traditionally studied adaptation via new mutations driven to fixation by natural selection and reducing genetic diversity on linked sites in a selective sweep (Smith & Haigh 1974). However, standing variation also plays an important role in adaptation (Barrett & Schluter 2008), and could result in a soft selective sweep when multiple haplotypes containing the beneficial allele are driven to fixation (Hermisson & Pennings 2005). Empirical studies have also found both kinds of genetic variants underlying adaptation, such as a new mutation in the gene *Agouti* underlying coat color adaptation in deer mice (Linnen *et al.* 2009) and evidence of genomewide standing variation underlying adaptation in songbirds (Lai *et al.* 2018). Selection on standing variation can also increase the fixation probability of weaker alleles, which may persist in relatively higher pre-adaptation frequencies than stronger effect alleles if the adaptive phenotype was deleterious prior to being favored (Hermisson & Pennings 2005). The presence of migration, on the other hand, can favor alleles with large effects that can withstand the homogenizing force of gene flow (Griswold 2006, Yeaman & Whitlock 2011), in some cases leading to the evolution of clusters of tightly linked genes affecting the trait (e.g. Jay *et al.* 2018, Matschiner *et al.* 2022).

Not surprisingly, empirical evidence shows that adaptive genetic architecture varies among species and among traits (Yeaman & Whitlock 2011). For example, polygenic adaptation with many small-effect loci underlies various human phenotypes (Pritchard & Di Rienzo 2010, Rockman 2012), while a large-effect locus can explain reduced body armor in sticklebacks (Schluter *et al*. 2021). An example of variation among species of the same genus includes mimicry patterns in *Heliconius* butterflies, which have a multilocus architecture with loci of different effect sizes for several species, but one, *H. lunata*, shows an architecture with a single large-effect locus (Baxter *et al*. 2009, Huber *et al*. 2015). Intraspecific variation can be seen in the genetic architecture of flowering time in *Arabidopsis thaliana*, which varies among different populations, showing that even when the same loci are involved their effect size can differ (Lopez-Arboleda *et al*. 2021). A variable architecture of adaptive melanism was also found among different different populations of *Drosophila melanogaster* (Bastide *et al*. 2016). Perhaps surprisingly, the genetic architecture of melanism in *D. melanogaster* was also variable among individuals within the same population (Bastide *et al*. 2016), and similarly polymorphic architectures were also inferred for ethanol resistance (Sprengelmeyer *et al*. 2021) and thorax and wing length (Sprengelmeyer *et al*. 2022). However, our understanding of the degree of genetic variability in ancestral and evolved

populations underlying adaptive trait differences is limited by the scale of prior studies.

To study this question, we expanded on the work of Bastide and colleagues (2016) on the evolution of melanic traits in high altitude *D. melanogaster*. *Drosophila melanogaster* is a cosmopolitan species that primarily lives alongside human settlements and has expanded out of its ancestral environment in southern Africa to colonize many different habitats approximately 13 kya (Sprengelmeyer *et al*. 2020). Coupled with its experimental tractability and the vast work on genetics and molecular biology accumulated for the species in the last century, *D. melanogaster* provides an important model to determine the genetic basis of local adaptation.

Drosophila melanogaster pigmentation, in particular, entails complex traits with a number of known genetic contributors (Rogers *et al*. 2014, Dembeck *et al*. 2015, Massey & Wittkopp 2016). Cuticular pigmentation synthesis involves dopa and dopamine, and the pathway includes the genes *ebony*, *tan*, and *yellow* (Wittkopp *et al*. 2003). In addition to how much melanin is produced, flies can also become darker by modulating where the pigment is allocated. In *D. melanogaster*, abdominal pigmentation pattern varies between males and females and is affected by expression patterns of the genes that regulate pigmentation synthesis, such as *bi*/*omb* and *bab* (Wittkopp *et al*. 2003). Evolutionary changes in gene expression of

pigmentation-related genes, beyond the few genes listed here, have recurrently been found to underlie variation in pigmentation between and within *Drosophila* species (Wittkopp *et al*. 2003, Massey & Wittkopp 2016). In many cases, the same genes seem to affect variation within and between species, as well as multiple pigmentation-related traits (Massey & Wittkopp 2016). Some of these same genes are also associated with adaptive melanism in high-elevation *D. melanogaster* populations (Pool & Aquadro 2007, Bastide *et al*. 2016), with an adaptive haplotype at the gene *ebony* shown to contain multiple cis-regulatory changes (Rebeiz *et al*. 2009).

Despite the existence of key genes underlying pigmentation, Bastide and colleagues still described a variable genetic architecture between and within populations, and no single quantitative trait locus (QTL) was detected across all mapping crosses (Bastide *et al*. 2016). This study also reported population genetic signals consistent with selection on standing variation at *ebony* and *tan*. These results suggested that the total pool of standing variation for pigmentation in a high altitude population may have been larger than what is necessary for each individual fly to become dark, and thus pigmentation traits may have reached their new optima before any specific favored variants reached fixation.

Herein, we aim to further investigate the variability of melanism adaptation genetic architecture within a population, and particularly to assess the degree of variability in ancestral range versus phenotypically

evolved populations. Among the populations studied by Bastide *et al*. (2016), we selected the high-elevation population from Ethiopia, which is the darkest known population of this species (Bastide *et al*. 2014). Our specific goals are to address the variability of the genetic architecture underlying adaptive melanism and answer the following questions: How much do QTLs vary as a function of the Ethiopian and Zambian (ancestral range) strains used in each cross? How similar are the genetic architectures of distinct melanic traits that have each evolved in Ethiopia? What are the potential genes and the effect sizes underlying common and uncommon QTLs? Lastly, considering that *D. melanogaster* pigmentation has a plastic response to temperature (David *et al*. 1990), we also ask whether the environment (cold *versus* warm) affects the presence and magnitude of QTLs in each cross.

To fulfill this aim, we expanded Bastide and colleagues' (2016) design by focusing more deeply on the architecture of pigmentation within a single melanic population (Ethiopia), analyzing 21 mapping crosses representing a grid of seven inbred dark Ethiopian strains crossed to three inbred light Zambian strains. We measured three pigmentation-related traits for all crosses instead of one trait per cross, and improved power by increasing the number of phenotyped female flies from 600 to 2,400 per cross. This greater sample size may also improve QTL resolution, alongside our increase in the number of generations from twenty to

twenty-six to allow for more recombination events. Lastly, for four of our crosses, we also measured flies raised at a derived-like 15 °C environment in addition to 25 °C. This greatly expanded mapping study allowed us to gain multiple insights into the genetic architecture of this model adaptive trait.

Methods

Mapping cross design

We used inversion-free inbred lines to generate mapping crosses between a dark population from the Ethiopian highland and a light population from Zambia, a population pair that has shown a variable genetic architecture for adaptive pigmentation in a previous study with four mapping crosses (Bastide *et al*. 2016). The Ethiopian population has the darkest individuals of *Drosophila melanogaster* that have been recorded (Bastide *et al*. 2014). The Zambian population is located in the species' ancestral range, in southern-central Africa (Sprengelmeyer *et al*. 2020). Additionally, these populations split around 13,000 years ago, and the Ethiopian population colonized the highland environment around 2,000 years ago (Sprengelmeyer *et al*. 2020), representing a relatively recent instance of local adaptation. The Ethiopian line was collected in Fiche (9.81°N, 38.63°E, altitude 3070 m), and the Zambian lines were collected in Siavonga (16.54°S, 28.72°E, altitude 530 m; Lack *et al*. 2015, Bastide *et al*. 2016).

We used seven dark inbred lines from Ethiopia (here coded as E1-E7) and three light inbred lines from Zambia (here Z1-Z3) (Table S1). We crossed each Ethiopian line to each Zambian line, resulting in 21 mapping crosses. Each started with a reciprocal cross of eight males from one inbred line and eight females from the other. Then, we combined 100 F1 females and 100 F1 males from each direction into a population of 400 flies in a single cage, where they were allowed to mate without generation overlap, at room temperature, until F25. We reared the F25 adults at 25 °C degrees with a 12:12 hour light:dark cycle, and sorted the F26 adults between 4 and 7 days old by phenotype.

Pigmentation is known to be a temperature-dependent phenotype (David *et al*. 1990), so we also analyzed phenotypes of four of the mapping crosses in a lower-temperature environment (to roughly approximate the cooler high altitude Ethiopian environment) one generation later. The crosses chosen used the Ethiopian inbred lines E4 and E7 and the Zambian inbred lines Z1 and Z3 as parental strains (Table S1). We moved F26 adults into bottles and reared them at 15 °C with a 12:12 hour light:dark cycle. Then, we sorted F27 adults between 4 and 7 days old by phenotype-color. We note that raising large numbers of flies at lower temperatures is considerably more demanding in time and effort; hence the reduced scope of our mapping effort in this environment.

Pigmentation phenotypes

We mapped three distinct traits per cross, sorting the same set of flies based on three commonly-studied pigmentation traits: the presence and intensity of the trident pattern on the dorsal thorax (hereafter, Trident), the pigmentation of the background of the fourth abdominal segment (hereafter, A4 Background), and the proportion of the fourth abdominal segment that was covered by the black stripe (hereafter, Stripe). Relatively low correlations among these three traits were observed among independent isofemale strains within each population ($r < 0.34$), except for a higher correlation between Trident and A4 Background within Ethiopia ($r = 0.49$; Bastide *et al*. 2014).

We sorted ~2400 female flies from each cross, except for two crosses in which only ~1600 female flies were obtained (Z3E3 at 25 °C and Z1E4 at 15 °C). The flies were combinatorically separated into the 5% darkest and lightest extremes for each trait. In a subset of the crosses, more than 5% of the flies had no trident, and all the flies without the trident were grouped together in these cases.

Due to the correlation among the traits, some flies had a phenotype in the 5% extreme of multiple traits. As expected from the correlations, elevated trait overlap was always in the same direction: 5% darkest for multiple traits, or 5% lightest for multiple traits, never 5% darkest for one trait and 5% lightest for another. Therefore, for each cross, we obtained

seven pools of the darkest flies and seven pools of the lightest flies: three pools of flies that were outliers for just one trait, three pools that were outliers for two of these traits but not the third, and one pool of flies that were outliers for all three traits.

DNA Extraction and Sequencing

The genome sequences of most parental strains were obtained from previous studies (Lack *et al*. 2016, Sprengelmeyer & Pool 2021, see Table S1), and for the sequences of three additional strains (E3, Z1, and Z2, see Table S1) we extracted DNA from 30 female flies and followed the protocol described in this section. For each phenotype pool, we extracted DNA in groups of up to 30 female flies using a protocol modified from (Lack *et al*. 2015). Instead of a phenol:chloroform extraction, crude DNA lysate was purified using a 1X SPRI bead cleanup (Sera-Mag Magnetic SpeedBeads, GE Healthcare). SPRI beads were prepared in a buffer of 10 mM Tris base, 1 mM EDTA, 2.5 M NaCl, 20% PEG 8000, 0.05% Tween 20. DNA concentrations were determined using Qubit (ThermoScientific), and equal quantities of DNA per fly were combined to create a DNA pool for each extreme phenotype. This process involved (1) combining individual extractions from flies exhibiting the same extreme trait but extracted using multiple tubes (e.g., if the 5% darkest A4 Background pool had 90 flies, three separate tubes would be used), and (2) combining extractions from

pools comprising flies exhibiting more than one extreme trait (e.g., flies that were in the 5% extreme for both A4 Background and Stripe were included in the DNA mixtures for both of those traits). Following the integration of individual extractions, a single DNA pool was obtained for each extreme trait per cross, resulting in six DNA pools per cross.

We prepared genomic libraries following Adams and colleagues (2020), with a left-handed size selection cutoff at 300 bp, and performed a cleanup using the Zymo Select-a-Size DNA clean & Concentrator MagBead Kit. Libraries were sequenced at the UW-Madison Biotechnology Center on the Illumina NovaSeq 6000 platform, with paired-end reads of 150 bp.

DNA Alignment

We mapped the pair-ended reads to the *D. melanogaster* (v5.57) reference genome using BWA-MEM, version 0.7.17 (Li 2013). We used samtools v1.13 to remove duplicate reads (Li *et al.* 2009) and GATK v3.2.2 to realign the around small indels using IndelRealigner (DePristo *et al.* 2011). For the parental genomes, we performed an additional step with GATK to generate vcf files with all the sites in the genome using UnifiedGenotyper and a base quality threshold equal to 10.

Quantitative trait loci (QTL) mapping

We detected QTLs using Simulation-based Inference for Bulk Segregant Analysis Mapping (SIBSAM, Pool 2016), which has the ability to

detect multiple linked QTLs in the same chromosomal region. First, we estimated "ancestry difference" (here, the difference in the proportion of local ancestry from the Ethiopian parental strain between the dark outlier pool and the light outlier pool) across every window in the genome. We defined window sizes to contain 250 non-singleton SNPs in the Zambian population, resulting in windows of ~8.5 kb and matching the ones used by Bastide and colleagues (2016). For each window, we counted the number of reads with information to distinguish one ancestry from another (informative reads). Then, we calculated the ancestry difference (a_d) as the proportion of Ethiopian reads among the total informative reads from the dark pool minus the proportion of Ethiopian reads among the total informative reads from the light pool.

We used SIBSAM's standard methods to identify the QTLs and their effect sizes (Pool 2016). In short, this method uses ancestry difference as a summary statistic to delineate QTLs, and implements a workflow consisting of (1) using null simulations (no real QTLs) to determine the significance of primary QTL peaks, then (2) using single QTL simulations to estimate effect sizes and confidence intervals of primary QTL peaks, along with the significance of any secondary QTL peaks associated with each primary QTL, and (3) using multi-QTL simulations to jointly estimate the effect sizes and confidence intervals of all significant QTLs in a given cluster of a primary and one or more secondary QTL peaks. To be

conservative with respect to the amount of recombination in our experiment, estimated confidence intervals that were particularly short, with sizes smaller than 0.25 cM on either side of the QTL peak, were expanded to a minimum of 0.25 cM *a posteriori*, resulting in minimum QTL confidence interval sizes of 0.5 cM.

Comparing QTLs across experimental dimensions

To assess the variability of genetic architectures among the traits and the parental strains of each population, we initially focused on the results from all mapping experiments performed at 25 °C. Then, to investigate the effect of temperature, results from the subset of four crosses measured at both cold (15 °C) and warm (25 °C) temperatures were analyzed.

Each trait mapping experiment at 25 °C has three variables: Ethiopian parental strain, Zambian parental strain, and pigmentation trait mapped. To investigate the effect of one of these variables on the probability of finding a QTL in a different mapping experiment, we compared these experiments while varying one variable at a time. For a given focal variable, we counted how often a QTL was found (*i.e.* overlapped at least one QTL) in another mapping experiment that had the other two variables fixed, and divided that total by the number of QTLs tested. This QTL overlap proportion metric varies from zero, if no QTL is found in any other compared mapping experiment, to 1, if each QTLs is found in all of

the compared experiments. Note that when analyzing only the 25 °C crosses, to account for the uneven sample size of Ethiopia strains (seven Ethiopian strains, *versus* three Zambian strains and three pigmentation traits), we downsampled the Ethiopian strains to groups of three and repeated the analysis for all 35 possible combinations of three Ethiopian strains and averaged the final scores. For the analysis including both cold and warm environments, which uses two Ethiopian strains, two Zambian strains, two temperatures, and three pigmentation phenotypes, we downsampled pigmentation to all possible combinations of two phenotypes. Ultimately, the score in this test indicates the likelihood of finding shared QTLs within the fixed variables and within and between the non-fixed variables, which sheds light on which variables are the strongest determinants of genetic architecture in our data set.

Additionally, for the cold and warm analysis, we compared the genome-wide pattern of ancestry differences (a_d) between the mappings performed at different temperatures. We calculated Δa_d as a_d in the cold experiment minus the warm experiment. Large a_d indicates a larger proportion of Ethiopian (dark parental population) alleles in the dark offspring pool than in the light pool. Large Δa_d positive numbers indicate that this effect was more pronounced in the cold than in the warm experiment, therefore suggesting a larger phenotypic effect size in the cold experiment. To focus on regions of extreme differences, we also

investigated which pigmentation genes overlapped regions with Δa_d higher than 0.30 or lower than -0.30.

Candidate genes and population genetics

To identify genes that might underly pigmentation evolution within our QTLs, we scanned the QTL regions for genes from a curated list of known pigmentation-related genes based on the FlyBase (release FB2023_06) "cuticle pigmentation" annotation (Gramates *et al*. 2022) and literature review (Rogers *et al*. 2014, Dembeck *et al*. 2015, Massey & Wittkopp 2016). Whether or not a given QTL contained any known pigmentation-related genes, we also identified genes that showed population genetic signatures of local adaptation between Ethiopia and Zambia. We used genome sequences from individual inversion-free inbred lines from Ethiopia, France, and Zambia to calculate all the three statistics as described in da Silva Ribeiro et al. (2022). We focused on QTL regions with recombination rates above 0.5 cM/Mb (Comeron *et al*. 2012) due to their more localized signatures of selection and we selected the genes associated with windows in the top 1% quantile of any of three population genetic statistics: the haplotype statistic χ_{MD} (Lange & Pool 2016) and two F_{ST} statistics (using Reynolds *et al*. 1983), $F_{ST\_FullWin}$ (window-wide F_{ST} calculated using all the SNPs in the window) and F_{ST_MaxSNP} (the highest F_{ST} value from a single SNP within the window). We included both the genes

that overlapped windows in the top 1% for a given chromosome arm, and the genes associated with the next nearest exon beyond each window boundary, to account for signatures of selection on cis-regulatory regions.

Results

QTLs vary by mapping cross and phenotype, but show striking overlap with known pigmentation genes

We mapped a total of 63 traits at 25 °C (three traits per cross for each of the 21 crosses involving seven Ethiopian and three Zambian inbred strains; Table S1). We sorted an average of 2,353 females per cross (Table S1), based on their combinatoric outlier status (top or bottom 5%) for thoracic trident intensity, abdominal background intensity, and abdominal stripe width (Materials and Methods). We obtained high depth of coverage data for all 126 extreme phenotype pools, averaging 39x depth per site (Table S2). We identified a total of 207 QTLs, ranging from zero (in ten cases) to 13 in each distinct trait mapping experiment (Figure 1, Figure S1-3, Table S2). The average maximum effect size was 0.223 across all 25 °C experiments with detected QTLs, with the greatest effect size observed for Z1E1-Trident (effect size = 0.414; Figure 2, Table S3).

The three most common QTLs overlapped the genes most canonically involved in pigmentation evolution: *yellow*, *tan*, and *ebony*. Out of the 207 QTLs, 74 overlapped at least one of these genes. Twenty-nine

mapping experiments had QTLs that overlapped *yellow* (left/telomeric end of chromosome X). The majority of these QTLs also included the pigmentation-related gene *Hormone receptor 4* (*Hr4*), but two mapping experiments had a narrower QTL that only overlapped *Hr4* and not *yellow*. Given the proximity to *yellow*, it is possible that *yellow* nevertheless contributes to these QTL signals but the confidence interval was shifted due to the effects of other linked pigmentation loci that were not distant enough to generate significant secondary QTL peaks. Twenty-six mapping experiments had QTLs that overlapped the gene *tan* (middle of chromosome X). Nineteen mapping experiments had QTLs that overlapped *ebony* (middle of chromosome 3R). An additional eight QTLs occurred within 100 kb of *ebony*, including the QTL with the strongest effect size (0.414) in the experiment. As suggested for *yellow*, it is possible that *ebony* is a primary driver of at least some of these nearby QTLs, with confidence intervals being shifted by the contributions of at least one other nearby pigmentation locus. In general, the QTLs with the strongest phenotypic effects overlapped *yellow*, *tan* and *ebony* (or nearly overlapped in the above-mentioned case), and QTLs in these regions tended to be stronger than those elsewhere (Figure 2, Table S3). These results are congruent with the prior expectation that these three major pigmentation genes played an important role in the adaptive evolution of melanism. However, it is also clear from the variable occurrence of these QTLs that

none of these genes individually is necessary for the evolution of dark Ethiopian flies. Most of the other QTLs with moderate effect size also overlapped pigmentation genes, such as *crol* (Rogers *et al*. 2014) on chromosome 2L, *jing* and *pdm3* (Rogers *et al*. 2014) on chromosome 2R, and *bab1* and *bab2* (Kopp *et al*. 2000) on chromosome 3L.

QTL presence is strongly dependent on both Ethiopian and Zambian parental strains

We conducted an analysis of shared (overlapping) QTLs between pairs of mapping experiments in which only one variable differed (either Ethiopian parent, Zambian parent, or trait). Here, our expectation is that for a more consequential variable (one that strongly influences QTL presence), QTL overlap will be relatively lower when that factor is varied; whereas, for a less consequential variable, QTL overlap will be relatively higher when it is varied. We found that the highest probability of finding a shared QTL occurred between different pigmentation phenotypes within the same parental cross. A QTL had a 41.62% probability of overlapping another QTL from the same parental strains, but a different trait, while the average null probability of that happening by chance was 19.9% and none of the 10,000 whole data permutations showed a value as extreme as 41.62%. Thus, in spite of the generally low correlations observed among

these traits within the Ethiopia and Zambia populations (Bastide *et al*. 2014), QTLs overlapped fairly often between traits for the same cross.

When only the Zambian strain was varied, the probability of a QTL being shared was 29.31% (p-value = 0.012 versus null permutations). A very similar QTL overlap proportion was observed when only the Ethiopian strain was varied (29.07%, p-value < 0.0001). Overall, these results shows that parental strain has a higher influence on the probability of two crosses sharing a QTL than the measured trait, suggesting that distinct parental strains carry distinct genetic variants underlying pigmentation. The influence of Zambian parental strains is congruent with selection on standing variation, but whereas such variants are often assumed to be rare prior to adaptation, the strong impact of within-Zambia variation in our study suggests that the favored variants occurred at appreciable frequencies in the ancestral population. The equally strong influence of Ethiopian parental strains, on the other hand, is congruent with partial sweeps, with selective variants not reaching fixation or very high frequency in the adapted population.

Candidate genes and signals of local adaptation within QTLs

Regarding the genes underlying adaptive pigmentation evolution, we found that 72% (150/207) of the QTLs contained at least one previously reported pigmentation gene, collectively encompassing 76% (99/131) of our

curated list of pigmentation genes. Of the remaining QTLs, twelve were in regions of low recombination in which signals of positive selection tend not to be well-localized, and 44% (20/45) of the remaining QTLs overlapped genomic regions with population genetic signatures of local adaptation (*i.e.* at least one genomic window with a given QTL was within the chromosome arm's top 1% of values for either window F_{ST}, maximum SNP F_{ST}, or the haplotype statistic χ_{MD} between Ethiopia and Zambia). It is worth mentioning that signatures of selection between the Ethiopian and Zambian populations also involve other adaptations besides pigmentation, and hence some local adaptation signals will reflect the evolution of unrelated traits, but at least one gene in each QTL is expected to be associated with pigmentation. In total, we have candidate genes or population genetic outliers associated with 82% (170/207) of our QTLs. An example can be seen in a QTL on chromosome X for Z2E3 A4 Background, which does not contain previously reported pigmentation genes but has three candidate genes under selection, *CG43287*, *CG15478*, and *CHES-1-like* (Figure 3, Table S3). The gene *CG15478* is an unannotated gene predicted to be a transcription factor (FlyBase, release FB2023_06), and could be associated with pigmentation. The gene *CHES-1-like* (*checkpoint suppressor 1-like*) has been shown to regulate the transcription of the gene *dpp* (*decapentaplegic,* not found in the QTL region) in the testis (Yu *et al*. 2016). The *dpp* signaling pathway controls abdominal stripe patterning (Wittkopp

et al. 2003), raising the question of whether *CHES-1-like* could also be regulating its expression in the cuticle and be involved in pigmentation evolution.

Another intriguing example is our strongest QTL (effect size = 0.414), a QTL for Z1E1 Trident that does not contain any pigmentation gene or signatures of selection but is located 73 kb away from *ebony* (Figure S4). If other genes affecting the trait are located too close to be considered a secondary QTL peak by SIBSAM, which assumes only one underlying locus affecting the QTL, their phenotypic effects could shift the position of the detected QTL. An inspection of the region does reveal genes such as *dmrt93B* (*doublesex-Mab related 93B*) and *r-l* (*rudimentary-like*), both overlapping the QTL peak. The related gene *doublesex* (*dsx*, not within this region) has been shown to regulate the expression of *yellow*, *tan*, and *ebony* (Massey & Wittkopp 2016). The gene *rudimentary* (also not within this region) is involved in *de novo* pyrimidine biosynthesis, and has also been associated with pigmentation, mainly through regulation of the gene *black* and the production of β-alanine instead of melanin (Piškur *et al*. 1993, Rawls Jr., 2006). The gene *suppressor of rudimentary* (*su(r)*), also involved in the β-Alanine pigmentation pathway, was included in our list of pigmentation genes (Dembeck *et al*. 2016). Whether the candidate genes under selection detected in our QTLs are also related to pigmentation needs further research.

Temperature effect on pigmentation QTL polymorphism

Among the four crosses studied at 15 °C in addition to 25 °C (involving all pairings among two Ethiopian and two Zambian strains), we found 42 and 43 total QTLs from the cold and warm treatments, respectively (Figure 4, Table S3). Most of these QTLs were found in both treatments, but 10 QTLs were exclusive to 15 °C and 6 to 25 °C. The overall effect sizes distribution was also similar (Figure S5, Table S3), with mean effect sizes equal to 0.167 and 0.137 in the cold and warm environments, respectively (t-test p-value = 0.081). These results are corroborated by our quantitative analysis of QTL overlap, mirroring the one described above for 25 °C, in which one of the cross-design factors (Ethiopian strain, Zambian strain, pigmentation trait, and now also temperature treatment) is allowed to vary while the other remains fixed. The probability of finding a shared QTL when temperature is the trait allowed to vary was 53.94%. When temperature is held fixed, the probabilities were 33.7%, 29.8%, and 27.1% when the Zambia strain, Ethiopia strain, and pigmentation trait, respectively, were allowed to vary. These results suggest that temperature treatment has a lower influence on whether a QTL can be detected in more than one cross than the other factors. Although these results point to a fairly similar genetic architecture at both temperatures, there are still important differences. For example, Z1E4 Stripe and Z3E4 A4 Background

only have a QTL overlapping *yellow* when raised at 15 °C, and no other trait mapping involving E4 has a QTL overlapping *yellow*. Although this result is restricted to the E4 parental strain, it is congruent with findings that cis-regulation of *yellow* is sensitive to temperature (Gibert *et al*. 2017). Of these QTLs, only the A4 background QTL extends to also include *crm*, which has a known role in the thermal plasticity of pigmentation (Gibert *et al*. 2011).

Additionally, when we compared the genomic landscapes of ancestry difference between temperatures, based on the deviation between 15 °C and 25 °C in this quantity (*i.e.* Δa_d), we saw that there are multiple regions with apparent signals of thermal plasticity (Figure 5). We see instances, for example, near the locations of *tan* and *ebony*, where the ancestry difference was higher in the cold than in the warm experiments, particularly for A4 Background and Trident. For both instances in which *tan* shows higher ancestry difference in the cold, plasticity seemed to depend on the Zambia strain involved in the cross, whereas, for the instance in which *ebony* shows higher ancestry difference in the cold, the signal was specific to one pair of parental strains. Results for both of these loci are congruent with a scenario in which colder temperatures enhance the effects of adaptive pigmentation variants. There are also regions showing the opposite pattern, for example overlapping the genes *bab1* and *bab2* in the Z3 × E4 cross, in which the effect of the Ethiopian alleles on A4 Background was higher in the warm temperatures, which could hinder selection on this

specific trait in a cooler environment. Other genes within strong QTL effects, such as *yellow*, *crol*, *jing*, and *pdm3*, did not show notable deviations in Δad, suggesting that their contributions to Ethiopia-Zambia pigmentation differences do not depend on the temperature.

In some cases, the same genomic regions in the same cross had extremely high Δa_d for one or two traits and extremely low Δa_d for the others. For example, the region around *tan* showed a strong positive Δa_d for trident and A4 background, but a strong negative Δa_d for stripe width. Similarly, the *ebony* region showed strong positive Δa_d for A4 background in one cross, but a strongly negative signal for stripe width. A region on 3L overlapping the *bab1/bab2* genes showed similar complexity: when E4 was paired with Z1, all three traits showed enhanced ancestry difference at 15 °C, but when E4 was instead paired with Z3, A4 background showed a strongly reduced signal at 15 °C. Hence, there may be complex relationships between plasticity and pleiotropy that modulate adaptive phenotypes and thus the trajectories of potentially favored variants.

Discussion

An instance of adaptive trait evolution in nature may involve changes in anywhere from one to numerous genes. On the one hand, polygenic adaptation has been proposed to involve standing variation at many loci with individually weak phenotypic effects and subtle allele frequency

changes that do not produce a selective sweep signature (Pritchard *et al*. 2010, Pritchard & Di Rienzo 2010). On the other hand, genetically simple instances of adaptive trait evolution are also known, such as the evolution of pigmentation in peppered moths (van't Hof *et al*. 2011) and DDT resistance in *Drosophila melanogaster* (Daborn *et al*. 2002). Recent theoretical work has shown that adaptation can also proceed in between these extremes of polygenic adaptation and complete selective sweeps, affecting several strong effect loci with moderate frequency changes and producing partial sweeps (Höllinger *et al*. 2023). Especially when there is ample standing genetic variation for selection to act upon, genetic redundancy underlying adaptive traits could result in the selection of several strong effect loci as well, as shown by experimental evolution in *Drosophila simulans* (Barghi *et al*. 2019). As with earlier QTL mapping studies on multiple adaptive traits in *D. melanogaster* (Bastide *et al*. 2016; Sprengelmayer & Pool 2021, Sprengelmeyer *et al*. 2022), our results indicated a persistently variable genetic basis for evolved traits, in which different individuals in the derived population have distinct genetic architectures underlying similar phenotypes, implying partial selective sweeps of favored variants. By independently varying the parental strains from both populations in a controlled manner, we also provide clear evidence that selection acted upon common standing variation. Notably then, even for pigmentation traits that are thought to have relatively

simpler molecular underpinnings than many complex traits, adaptive evolution in this genetically diverse species was apparently not limited by available variation; and potentially due to this hypothesized excess of melanic variants available to selection, each individual only requires a subset of the adaptive variants to become darker than the ancestral population, and none of them have reached a frequency approaching fixation.

These results are congruent with Bastide and colleagues (2016), who studied melanic evolution in three highland populations, including Ethiopia. Comparing our results, we found that within-population variation was similar to the variation found across their three different populations. Out of the nineteen QTLs detected by them, only one did not overlap QTLs from our study: Q12, a QTL restricted to Cameroon. As the different populations also show similar variability with each other, we argue that the variation in genetic architecture we see within the Ethiopian population is the rule for this trait across other populations as well. This is congruent with the hypothesis that common beneficial standing variation exists within the ancestral range and that different derived populations were founded by individuals carrying polymorphisms in many pigmentation loci (in spite of the mild population bottleneck that may have accompanied the species' expansion into the Ethiopian highlands; Sprengelmeyer *et al*. 2020).

The presence of strong QTLs in some mapping experiments but not others is unlikely to be due to lack of power. According to SIBSAM's power analysis, we would have nearly 100% probability of detecting QTLs of effect size 0.15 or greater. While persistent polymorphism is the simplest explanation for these results, an additional possibility is that pigmentation QTLs can be dependent on the combination of parental strains due to epistasis. Using a recombinant inbred line panel from one of the same Ethiopia-Zambia crosses studied here, we did not find evidence of second-order epistasis for pigmentation traits (Chapter 3), but given the variability observed among our mapping experiments, it is possible that either the use of different crosses or the investigation of higher-order epistasis could show different results.

A simpler explanation for the absence of QTL underlying pigmentation differentiation, however, is that either the dark parental strain did not have the adaptive allele for that locus (congruent with our proposed non-fixation of adaptive alleles), or that the light parental strain did have the adaptive allele (congruent with the proposed common standing variation of adaptive alleles). The dark allele could be present in the ancestral population without resulting in a dark phenotype if its effect size depended on genetic background, for example. Further investigation of the expression pattern of pigmentation genes in the different parental

strains could be a helpful approach to distinguish between these explanations.

Functional analysis could also assist in validating the role of genes with signatures of selection detected in our QTLs, as well as how many genes actually underlie each QTL. It is hard to determine whether one or multiple genes underlie a QTL, and few studies have managed to dissect the exact gene or genes underlying a QTL. In a study of maize domestication, a strong QTL overlapping the gene *teosinte branched1* (*tb1*) has been found in multiple traits, and interestingly, for one trait *tb1* was the sole gene underlying the QTL, but for others, additional genes were detected on the QTL region, including genes that interacted epistatically with *tb1* (Studer & Doebley 2011).

Regarding the genetic architecture among traits, the shared QTLs suggest a partially pleiotropic genetic basis, despite our choice of pigmentation traits with low or at most moderate correlations within Ethiopia and Zambia (Bastide *et al*. 2014). Many genes in the cuticle pigmentation pathway in *D. melanogaster* have been shown to affect distinct pigmentation traits, such as the effect of *yellow*, *ebony,* and *tan* on both thorax and abdominal pigmentation, although tissue-specific enhancers are known to exist (Wittkopp *et al*. 2002, Massey & Wittkopp 2016, Endler *et al*. 2018). The phenotypic effect on multiple traits was stronger among the twenty-one crosses measured at 25 °C than in the analysis using only the

four crosses measured at 25 °C and 15 °C, in which we saw that some loci responded in the opposite directions in different temperatures (Figure 5). Therefore, while moving into cooler environments might have revealed variation for some traits for selection to act upon, it might have hindered selection on other traits.

Overall, our greatly expanded mapping study recapitulates inferences of a polymorphic genetic architecture of adaptive melanism in flies (Bastide *et al*. 2016). But importantly, we showed here that melanic evolution involves the persistence of common polymorphisms at causative variants not only in the derived population, but also in the ancestral population. Our results support a model of adaptation involving partial sweeps on common standing variation, congruent with recent theoretical (Höllinger *et al*. 2023) and experimental studies (Barghi *et al*. 2019). Further investigation of the expression pattern of major pigmentation genes such as *yellow*, *tan*, and *ebony*, in both Ethiopian and Zambian inbred strains could be helpful to validate our findings, as well as further functional dissection of natural alleles at genes potentially underlying strong QTL regions.

Acknowledgements

We would like to thank Clarice Danen for the assistance with fly husbandry. I also would like to thank NIH grants R01 GM127480 and R35 GM13630 to JEP.

References

Adams, M., McBroome, J., Maurer, N., Pepper-Tunick, E., Saremi, N. F., Green, R. E., Vollmers, C., & Corbett-Detig, R. B. (2020). One fly–one genome: chromosome-scale genome assembly of a single outbred *Drosophila melanogaster*. *Nucleic Acids Research*, 48(13), e75-e75. https://doi.org/10.1093/nar/gkaa450.

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS biology*, 17(2), e3000128. https://doi.org/10.1371/journal.pbio.3000128.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38-44. https://doi.org/10.1016/j.tree.2007.09.008.

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., & Pool, J. E. (2016). A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics*, 204(3), 1307–1319. https://doi.org/10.1534/genetics.116.192492.

Bastide, H., Yassin, A., Johanning, E. J., & Pool, J. E. (2014). Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan Africa. *BMC Evolutionary Biology*, 14(1), 179. https://doi.org/10.1186/s12862-014-0179-y.

Baxter, S. W., Johnston, S. E., & Jiggins, C. D. (2009). Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. *Heredity*, 102(1), 57-65. https://doi.org/10.1038/hdy.2008.109.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genetics*, 8(10), e1002905. https://doi.org/10.1371/journal.pgen.1002905.

da Silva Ribeiro, T., Galván, J. A., & Pool, J. E. (2022). Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation. *Genome Biology and Evolution*, 14(10), evac143. https://doi.org/10.1093/gbe/evac143.

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., Feyereisen, R., Wilson, T.G., & Ffrench-Constant, R. H. (2002). A single *P450* allele associated with insecticide resistance in *Drosophila*. *Science*, 297(5590), 2253-2256. https://doi.org/10.1126/science.1074170.

David, J. R., Capy, P., & Gauthier, J.-P. (1990). Abdominal pigmentation and growth temperature in *Drosophila melanogaster*: similarities and differences in the norms of reaction of successive segments. *Journal of Evolutionary Biology*, 3(5–6), 429–445. https://doi.org/10.1046/j.1420-9101.1990.3050429.x.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491-498. https://doi.org/10.1038/ng.806.

Dembeck, L. M., Huang, W., Magwire, M. M., Lawrence, F., Lyman, R. F., & Mackay, T. F. (2015). Genetic architecture of abdominal pigmentation in *Drosophila melanogaster*. *PLoS Genetics*, 11(5), e1005163. https://doi.org/10.1371/journal.pgen.1005163.

Endler, L., Gibert, J. M., Nolte, V., & Schlötterer, C. (2018). Pleiotropic effects of regulatory variation in tan result in correlation of two pigmentation traits in *Drosophila melanogaster*. *Molecular Ecology*, 27(16), 3207-3218. https://doi.org/10.1111/mec.14781.

Fisher, R. A. (1930). The genetical theory of natural selection (pp. xiv, 272). Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Gibert, J. M., Karch, F., & Schlötterer, C. (2011). Segregating variation in the polycomb group gene *cramped* alters the effect of temperature on multiple traits. *PLoS Genetics*, 7(1), e1001280. https://doi.org/10.1371/journal.pgen.1001280.

Gibert, J. M., Mouchel-Vielh, E., & Peronnet, F. (2017). Modulation of yellow expression contributes to thermal plasticity of female

abdominal pigmentation in *Drosophila melanogaster*. *Scientific Reports*, 7(1), 43370. https://doi.org/10.1038/srep43370.

Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., Dos Santos, G., Goodman, J. L., Goutte-Gattat, D., Jenkins V. K., Kaufman, T., Larkin, A., Matthews, B. B., Millburn, G., & Strelets, V. B. (2022). FlyBase: a guided tour of highlighted features. *Genetics*, 220(4), iyac035. http://dx.doi.org/10.1093/genetics/iyac035.

Griswold, C. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, 96, 445-453. https://doi.org/10.1038/sj.hdy.6800822.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4), 2335-2352. https://doi.org/10.1534/genetics.104.036947.

Höllinger, I., Wölfl, B., & Hermisson, J. (2023). A theory of oligogenic adaptation of a quantitative trait. *Genetics*, 225(2), iyad139. https://doi.org/10.1093/genetics/iyad139.

Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R., & Mackay, T. F. (2020). Genotype by environment interaction for gene expression in *Drosophila melanogaster*. *Nature Communications*, 11(1), 5451. https://doi.org/10.1038/s41467-020-19131-y.

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W. O., & Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, 114(5), 515-524. https://doi.org/10.1038/hdy.2015.22.

Jay, P., Whibley, A., Frézal, L., de Cara, M. Á. R., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, 28(11), 1839-1845. https://doi.org/10.1016/j.cub.2018.04.072

Kopp, A., Duncan, I., & Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature*, 408(6812), 553-559. https://doi.org/10.1038/35046017.

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., Langley, C. H., & Pool, J. E. (2015). The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*, 199(4), 1229–1241. https://doi.org/10.1534/genetics.115.174664.

Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B., & Pool, J. E. (2016). A thousand fly genomes: an expanded *Drosophila* genome nexus. *Molecular Biology and Evolution*, 33(12), 3308-3313. https://doi.org/10.1093/molbev/msw195.

Lange, J. D., & Pool, J. E. (2016). A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. *Molecular Ecology*, 25(13), 3081-3100. https://doi.org/10.1111/mec.13671.

Lai, Y. T., Yeung, C. K., Omland, K. E., Pang, E. L., Hao, Y. U., Liao, B. Y., Cao, H. D., Zhang, B. W., Yeh, C. F., Hung, C. M., Hung, H. Y., Yang, M. Y., Liang, W., Hsu, Y. C., Yao, C. T., Dong, L., Lin, K., & Li, S. H. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, 116(6), 2152-2157. https://doi.org/10.1073/pnas.1813597116.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-Bio]. http://arxiv.org/abs/1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352.

Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, 325(5944), 1095-1098. https://doi.org/10.1126/science.1175826.

Lopez-Arboleda, W. A., Reinert, S., Nordborg, M., & Korte, A. (2021). Global genetic heterogeneity in adaptive traits. *Molecular Biology and Evolution*, 38(11), 4822-4831. https://doi.org/10.1093/molbev/msab208.

Massey, J. H., & Wittkopp, P. J. (2016). The genetic basis of pigmentation differences within and between *Drosophila* species. *Current Topics in Developmental Biology*, 119, 27-61. https://doi.org/10.1016/bs.ctdb.2016.03.004.

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, 6(4), 469-481. https://doi.org/10.1038/s41559-022-01661-x.

Piškur, J., Kolbak, D., Søndergaard, L., & Pedersen, M. B. (1993). The dominant mutation Suppressor of black indicates that de novo pyrimidine biosynthesis is involved in the *Drosophila* tan pigmentation pathway. *Molecular and General Genetics MGG*, 241, 335-340. https://doi.org/10.1007/BF00284686.

Pool, J. E. (2016). Genetic mapping by bulk segregant analysis in *Drosophila*: experimental design and simulation-based inference. *Genetics*, 204(3), 1295-1306. https://doi.org/10.1534/genetics.116.192484.

Pool, J. E., & Aquadro, C. F. (2007). The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Molecular Ecology*, 16(14), 2844-2851. https://doi.org/10.1111/j.1365-294X.2007.03324.x.

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation–not by sweeps alone. *Nature Reviews Genetics*, 11(10), 665-667. https://doi.org/10.1038/nrg2880.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4), R208-R215. https://doi.org/10.1016/j.cub.2009.11.055.

Rawls Jr, J. M. (2006). Analysis of pyrimidine catabolism in *Drosophila melanogaster* using epistatic interactions with mutations of pyrimidine biosynthesis and β-alanine metabolism. *Genetics*, 172(3), 1665-1674. https://doi.org/10.1534/genetics.105.052753.

Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F., & Carroll, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science*, 326(5960), 1663-1667. https://doi.org/10.1126/science.1178357.

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105(3), 767-779. https://doi.org/10.1093/genetics/105.3.767.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 66(1), 1-17. https://doi.org/10.1111/j.1558-5646.2011.01486.x.

Rogers, W. A., Grover, S., Stringer, S. J., Parks, J., Rebeiz, M., & Williams, T. M. (2014). A survey of the trans-regulatory landscape for *Drosophila melanogaster* abdominal pigmentation. *Developmental Biology*, 385(2), 417-432. https://doi.org/10.1016/j.ydbio.2013.11.013.

Schluter, D., Marchinko, K. B., Arnegard, M. E., Zhang, H., Brady, S. D., Jones, F. C., Bell, A. B., & Kingsley, D. M. (2021). Fitness maps to a

large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences*, 118(3), e1914889118. https://doi.org/10.1073/pnas.1914889118.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35. https://doi.org/10.1017/S0016672300014634.

Sprengelmeyer, Q. D., Lack, J. B., Braun, D. T., Monette, M. J., & Pool, J. E. (2022). The evolution of larger size in high-altitude *Drosophila melanogaster* has a variable genetic architecture. *G3*, 12(3), jkab454. https://doi.org/10.1093/g3journal/jkab454.

Sprengelmeyer, Q. D., Mansourian, S., Lange, J. D., Matute, D. R., Cooper, B. S., Jirle, E. V., Stensmyr, M. C., & Pool, J. E. (2020). Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Molecular Biology and Evolution*, 37(3), 627–638. https://doi.org/10.1093/molbev/msz271.

Sprengelmeyer, Q. D., & Pool, J. E. (2021). Ethanol resistance in *Drosophila melanogaster* has increased in parallel cold-adapted populations and shows a variable genetic architecture within and between populations. *Ecology and Evolution*, 11(21), 15364-15376. https://doi.org/10.1002/ece3.8228.

Studer, A. J., & Doebley, J. F. (2011). Do large effect QTL fractionate? A case study at the maize domestication QTL teosinte branched1. *Genetics*, 188(3), 673-681. https://doi.org/10.1534/genetics.111.126508.

van't Hof, A. E., Edmonds, N., Dalíková, M., Marec, F., & Saccheri, I. J. (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science*, 332(6032), 958-960. https://doi.org/10.1126/science.1203043.

Wittkopp, P. J., Carroll, S. B., & Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *TRENDS in Genetics*, 19(9), 495-504. https://doi.org/10.1016/S0168-9525(03)00194-X.

Wittkopp, P. J., True, J. R., & Carroll, S. B. (2002). Reciprocal functions of the *Drosophila yellow* and *ebony* proteins in the development and evolution of pigment patterns. *Development*, 129 (8), 1849–1858. https://doi.org/10.1242/dev.129.8.1849.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, 65(7), 1897-1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x.

Yu, J., Liu, Y., Lan, X., Wu, H., Wen, Y., Zhou, Z., Hu, Z., Sha, J., Guo, X., & Tong, C. (2016). *CHES-1-like*, the ortholog of a non-obstructive azoospermia-associated gene, blocks germline stem cell differentiation by upregulating *Dpp* expression in *Drosophila* testis. *Oncotarget*, 7(27), 42303. https://doi.org/10.18632%2Foncotarget.9789.

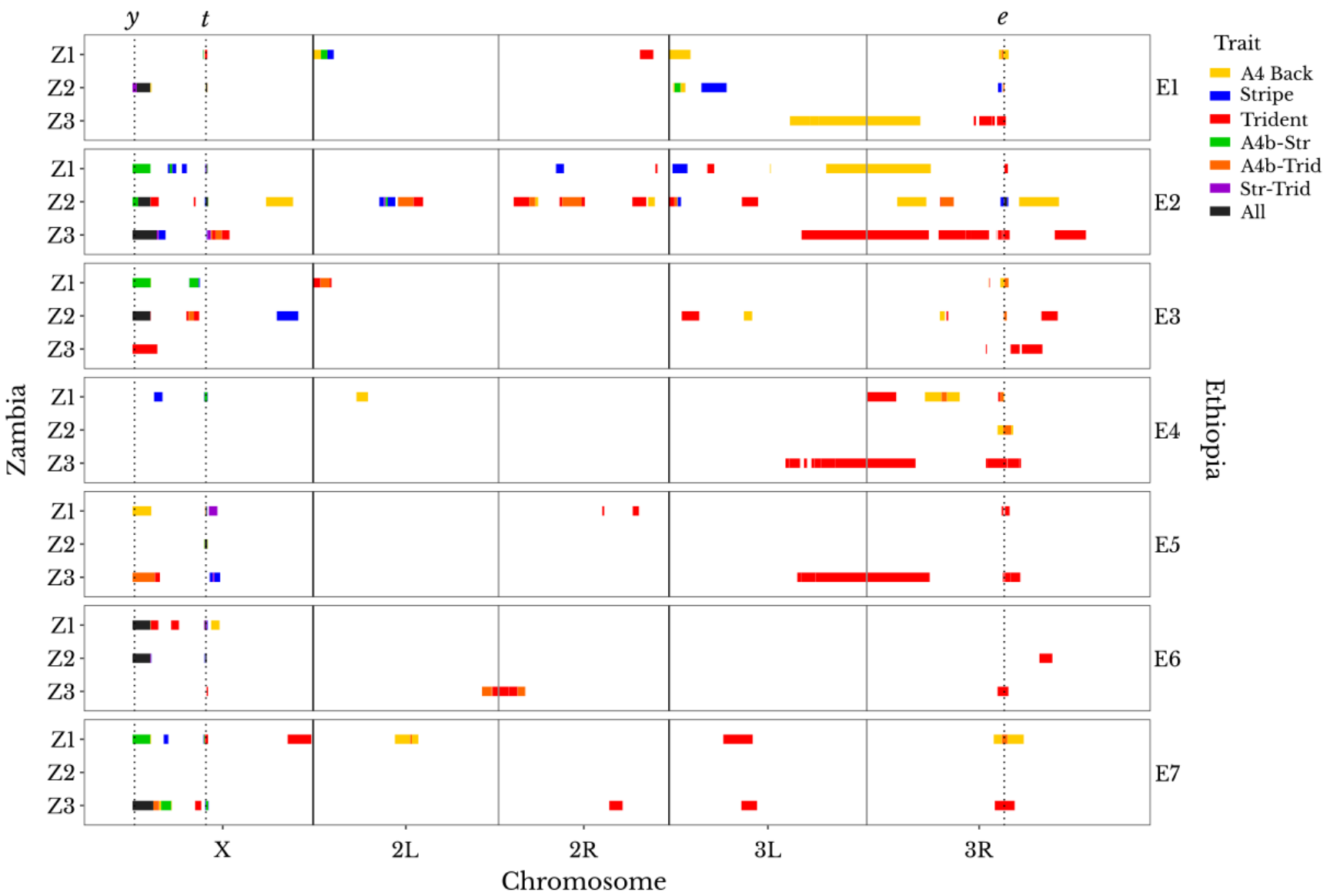

y
t
e
Trait
A4 Back
Stripe
Trident
A4b-Str
A4b-Trid
Str-Trid
All
Zambia
Ethiopia
Z1
Z2
Z3
E1
E2
E3
E4
E5
E6
E7
X
2L
2R
3L
3R
Chromosome

Figure 1. Variable genetic architecture of adaptive melanism in *Drosophila melanogaster*, with a high frequency of QTLs overlapping the canonical pigmentation genes *yellow*, *tan*, and *ebony*. Each row represents a mapping cross and contains the confidence interval of the QTLs detected for the mapped traits at 25 ˚C. Each plot block represents the Ethiopian parent (right y-axis), and the Zambian parent is shown on the left y-axis. The different traits are shown by different colors. Each primary color shows a region in which the QTL for a given trait did not overlap the QTLs for other traits in that mapping cross (yellow = A4 background, Blue = Stripe, Red = Trident). Each secondary color represents a region where QTLs for two traits overlap in that mapping cross (Green = A4 background and Stripe, Orange = A4 background and Trident, Purple = Stripe and Trident). Black represents regions where the QTLs for all three mapped traits overlapped in that mapping cross. Vertical black solid lines show the separation between chromosomes X, 2, and 3. Vertical gray lines show the separation between the left and right arms of chromosomes 2 and 3. Vertical dotted lines show the location of pigmentation genes (*y* = *yellow*, *t* = *tan*, *e* = *ebony*).

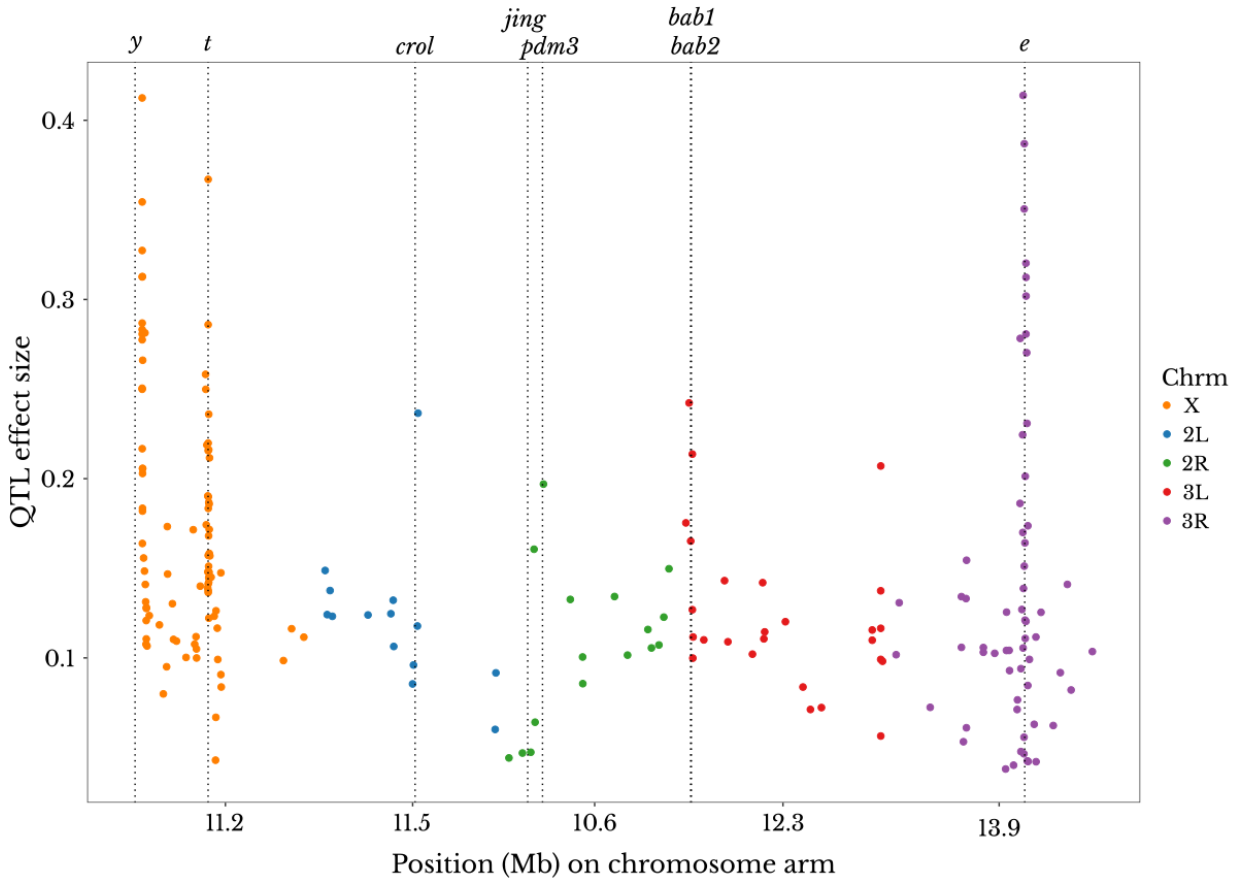


Figure 2. Distribution of the estimated effect sizes of QTLs underlying adaptive pigmentation in different mapping crosses, showing that the strongest effect QTLs overlap canonical pigmentation genes and other known regulators (labeled above). Colors represent different chromosome arms.

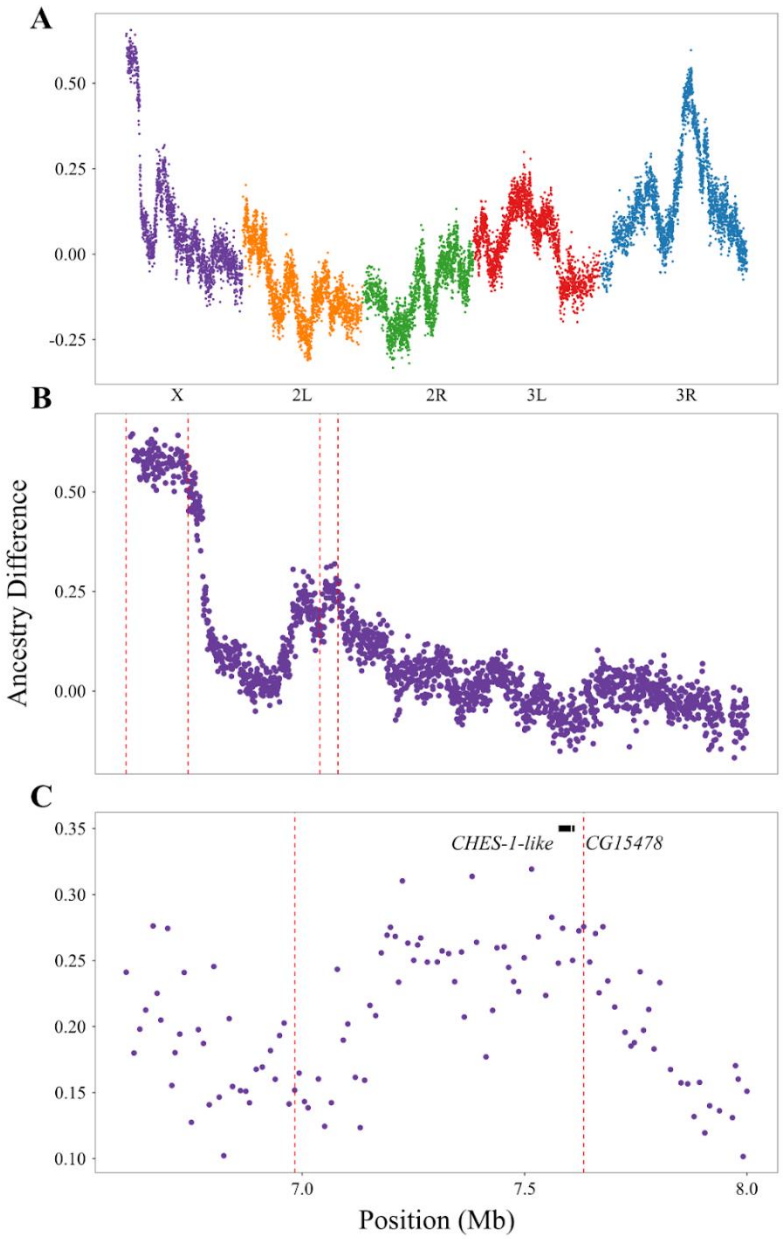


Figure 3. Ancestry distribution for trait mapping Z2E3 A4 Background shows two QTLs on chromosome X. (A) Genomewide ancestry difference distribution, different colors representing different chromosome arms. (B) Chromosome arm X. QTL confidence intervals shown in red dashed lines. (C) The second QTL on chromosome X, which did not include a gene

previously involved in pigmentation, showing candidate genes under selection detected with population genetics. QTL confidence interval shown in red dashed lines, *CHES-1-like* and *CG15478* shown as a black bar above the ancestry difference dots, gene *CG43287* not shown and located in an intronic region of *CHES-1-like*.

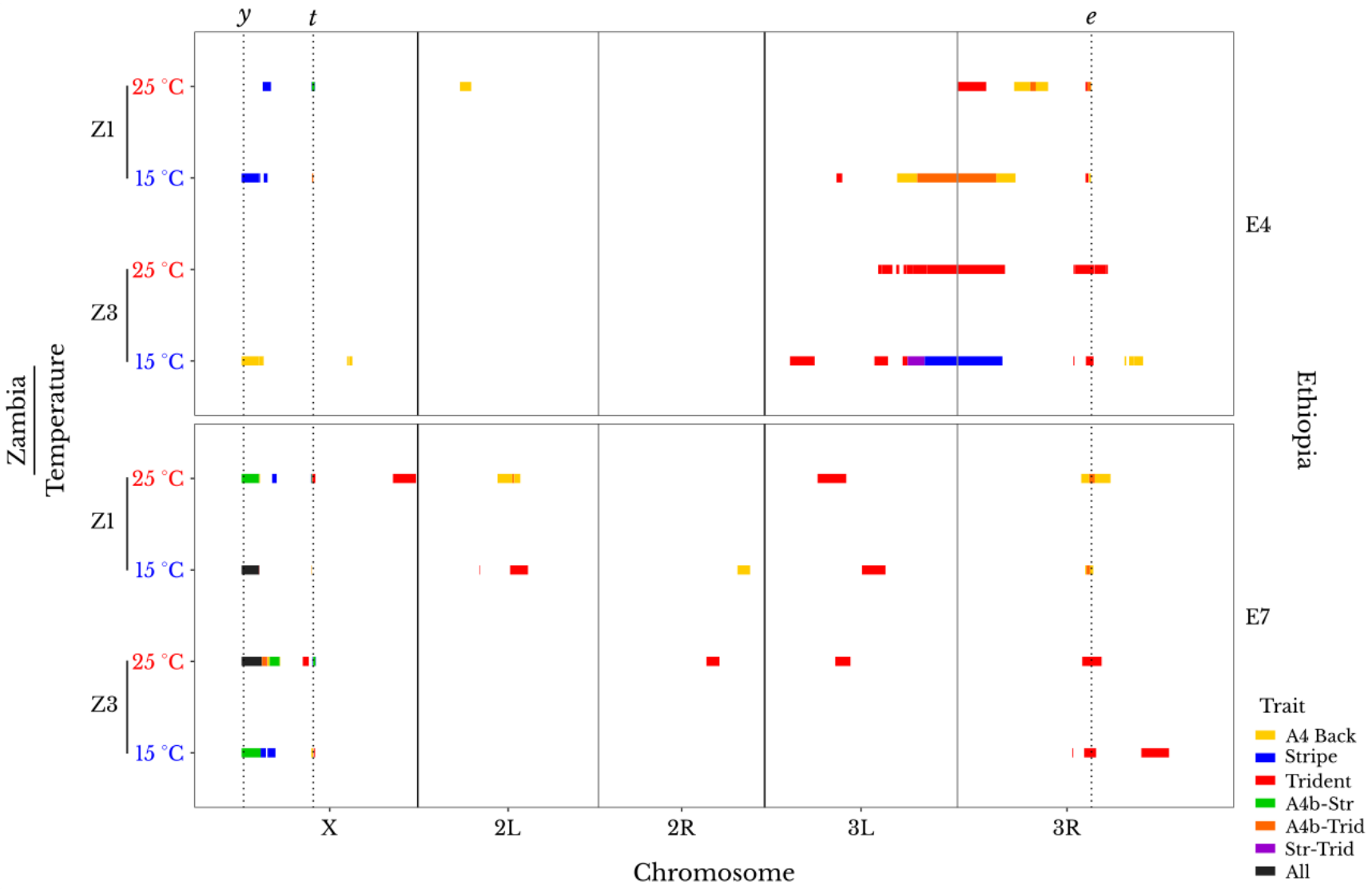

y
t
e
Zambia
Temperature
Z1
Z3
25 °C
15 °C
E4
E7
Ethiopia
X
2L
2R
3L
3R
Chromosome
Trait
A4 Back
Stripe
Trident
A4b-Str
A4b-Trid
Str-Trid
All

Figure 4. Significant QTLs are mostly similar at 15 °C and 25 °C, but show some differences. Each plot block represents the Ethiopian parent (right y-axis), and the Zambian parent is shown on the left y-axis. The y-axis also denotes the trait mapping temperature, showing the cold results directly below the warm results. The different traits are shown by different colors. As in Figure 1, each primary color shows a region in which the QTL for a given trait did not overlap the QTLs for other traits in that mapping cross (yellow = A4 background, Blue = Stripe, Red = Trident). Each secondary color represents a region where QTLs for two traits overlap in that mapping cross (Green = A4 background and Stripe, Orange = A4 background and Trident, Purple = Stripe and Trident). Black represents regions where the QTLs for all three mapped traits overlapped in that mapping cross. Vertical black solid lines show the separation between chromosomes X, 2, and 3. Vertical gray lines show the separation between the left and right arms of chromosomes 2 and 3. Vertical dotted lines show the location of pigmentation genes (*y* = *yellow*, *t* = *tan*, *e* = *ebony*).

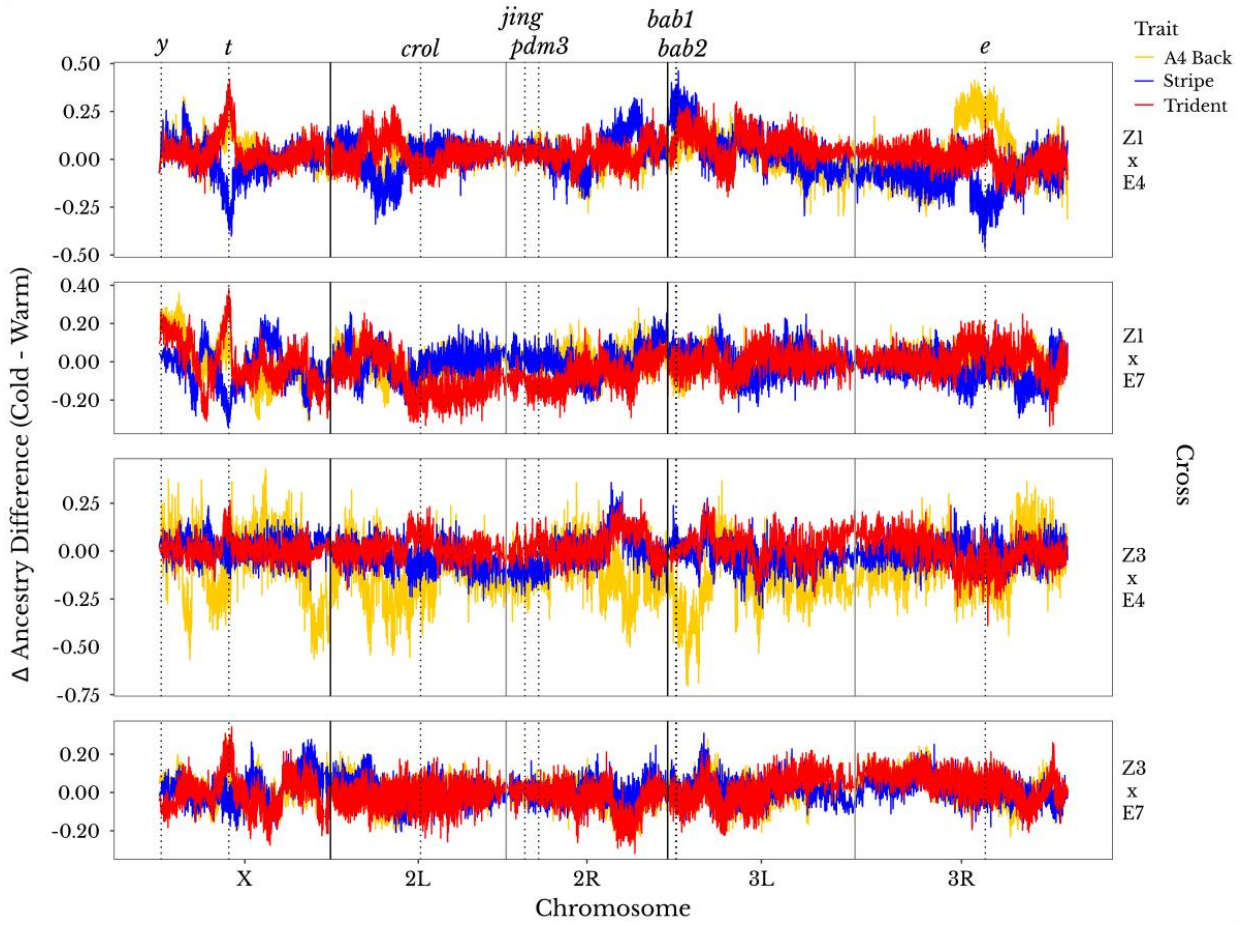


Figure 5. Genome-wide Δa_d (ancestry difference at 15 °C minus ancestry difference at 25 °C) illustrates signals of thermal plasticity in pigmentation differences. Δa_d is shown on the left y-axis. Each plot block represents the Ethiopian by Zambian cross (right y-axis). The different traits are shown by different colors (yellow = A4 background, Blue = Stripe, Red = Trident). Vertical black solid lines show the separation between chromosomes X, 2, and 3. Vertical gray lines show the separation between the left and right arms of chromosomes 2 and 3. Vertical dotted lines show the location of

pigmentation genes (*y* = *yellow*, *t* = *tan*, *crol* = *crooked legs*, *jing* = *jing*, *pdm3* = *pou domain motif 3*, *bab1* and *bab2* = *bric a brac 1* and *2*, and *e* = *ebony*).

Supplementary tables and figures

Table S1. Inbred lines used on the mapping crosses. Stock name refers to the access name used by the Pool Lab. Code refers to the code used within this article. 15 °C indicates whether the cross was used in the cold experiment. All crosses were used on the experiment at 25 °C.

Population	Stock Name	Code	15 °C	Sequence
Ethiopia	EF6N	E1		Lack *et al*. 2016
Ethiopia	EF9N	E2		Lack *et al*. 2016
Ethiopia	EF39N	E3		Sequenced here.
Ethiopia	EF43N	E4	Yes	Lack *et al*. 2016
Ethiopia	EF66N	E5		Lack *et al*. 2016
Ethiopia	EF95N	E6		Lack *et al*. 2016
Ethiopia	EF130N	E7	Yes	Lack *et al*. 2016
Zambia	ZI251N	Z1	Yes	Sequenced here.
Zambia	ZI413N	Z2		Sequenced here.
Zambia	ZI418N	Z3	Yes	Sprengelmeyer & Pool 2021

Table S2. Number of QTLs per trait mapping experiment. Numbers shown separately for each chromosome. Maximum and mean effect sizes for each experiment shown in the last two columns.

Temperature	Zambia	Ethiopia	Trait	QTLs				
				X	2	3	Max effect size	Mean effect size
15 °C	Z1	E4	A4 Background	1	0	2	0.371	0.252
15 °C	Z1	E4	Stripe	3	0	0	0.164	0.089
15 °C	Z1	E4	Trident	1	0	3	0.400	0.246
15 °C	Z1	E7	A4 Background	2	1	1	0.341	0.220
15 °C	Z1	E7	Stripe	1	0	0	0.396	0.396
15 °C	Z1	E7	Trident	2	2	2	0.330	0.184
15 °C	Z3	E4	A4 Background	4	0	3	0.172	0.092
15 °C	Z3	E4	Stripe	0	0	1	0.119	0.119
15 °C	Z3	E4	Trident	0	0	5	0.161	0.137
15 °C	Z3	E7	A4 Background	2	0	0	0.185	0.180
15 °C	Z3	E7	Stripe	2	0	0	0.158	0.137
15 °C	Z3	E7	Trident	1	0	3	0.243	0.140
25 °C	Z1	E1	A4 Background	1	1	2	0.250	0.162
25 °C	Z1	E1	Stripe	1	1	0	0.258	0.191
25 °C	Z1	E1	Trident	1	1	1	0.414	0.237
25 °C	Z1	E2	A4 Background	3	0	2	0.250	0.152
25 °C	Z1	E2	Stripe	4	1	1	0.287	0.152
25 °C	Z1	E2	Trident	1	1	2	0.270	0.195
25 °C	Z1	E3	A4 Background	2	1	2	0.182	0.131
25 °C	Z1	E3	Stripe	2	0	0	0.164	0.134
25 °C	Z1	E3	Trident	0	1	2	0.231	0.149
25 °C	Z1	E4	A4 Background	1	1	2	0.224	0.153
25 °C	Z1	E4	Stripe	2	0	0	0.141	0.130
25 °C	Z1	E4	Trident	0	0	4	0.186	0.135
25 °C	Z1	E5	A4 Background	2	0	0	0.216	0.211
25 °C	Z1	E5	Stripe	2	0	0	0.236	0.139
25 °C	Z1	E5	Trident	2	2	2	0.286	0.132
25 °C	Z1	E6	A4 Background	3	0	0	0.183	0.141
25 °C	Z1	E6	Stripe	2	0	0	0.217	0.177
25 °C	Z1	E6	Trident	3	0	0	0.148	0.126
25 °C	Z1	E7	A4 Background	2	1	1	0.219	0.155

25 °C	Z1	E7	Stripe	3	0	0	0.313	0.220
25 °C	Z1	E7	Trident	2	1	2	0.237	0.161
25 °C	Z2	E1	A4 Background	2	0	2	0.350	0.236
25 °C	Z2	E1	Stripe	2	0	3	0.281	0.196
25 °C	Z2	E1	Trident	2	0	1	0.387	0.253
25 °C	Z2	E2	A4 Background	3	5	5	0.302	0.150
25 °C	Z2	E2	Stripe	2	1	2	0.354	0.196
25 °C	Z2	E2	Trident	3	5	4	0.312	0.145
25 °C	Z2	E3	A4 Background	2	0	3	0.412	0.228
25 °C	Z2	E3	Stripe	2	0	0	0.327	0.222
25 °C	Z2	E3	Trident	2	0	4	0.320	0.181
25 °C	Z2	E4	A4 Background	0	0	1	0.120	0.120
25 °C	Z2	E4	Stripe	0	0	0	n/a	n/a
25 °C	Z2	E4	Trident	0	0	1	0.174	0.174
25 °C	Z2	E5	A4 Background	1	0	0	0.138	0.138
25 °C	Z2	E5	Stripe	1	0	0	0.148	0.148
25 °C	Z2	E5	Trident	1	0	0	0.183	0.183
25 °C	Z2	E6	A4 Background	2	0	0	0.278	0.249
25 °C	Z2	E6	Stripe	2	0	0	0.266	0.228
25 °C	Z2	E6	Trident	2	0	1	0.367	0.237
25 °C	Z2	E7	A4 Background	0	0	0	n/a	n/a
25 °C	Z2	E7	Stripe	0	0	0	n/a	n/a
25 °C	Z2	E7	Trident	0	0	0	n/a	n/a
25 °C	Z3	E1	A4 Background	0	0	2	0.116	0.113
25 °C	Z3	E1	Stripe	0	0	0	n/a	n/a
25 °C	Z3	E1	Trident	0	0	4	0.127	0.077
25 °C	Z3	E2	A4 Background	3	0	0	0.141	0.116
25 °C	Z3	E2	Stripe	3	0	0	0.145	0.112
25 °C	Z3	E2	Trident	3	0	9	0.147	0.090
25 °C	Z3	E3	A4 Background	0	0	0	n/a	n/a
25 °C	Z3	E3	Stripe	0	0	0	n/a	n/a
25 °C	Z3	E3	Trident	1	0	3	0.131	0.102
25 °C	Z3	E4	A4 Background	0	0	0	n/a	n/a

25 °C	Z3	E4	Stripe	0	0	0	n/a	n/a
25 °C	Z3	E4	Trident	0	0	10	0.207	0.100
25 °C	Z3	E5	A4 Background	1	0	0	0.148	0.148
25 °C	Z3	E5	Stripe	1	0	0	0.117	0.117
25 °C	Z3	E5	Trident	2	0	5	0.137	0.087
25 °C	Z3	E6	A4 Background	0	2	0	0.092	0.078
25 °C	Z3	E6	Stripe	0	0	0	n/a	n/a
25 °C	Z3	E6	Trident	1	4	1	0.212	0.091
25 °C	Z3	E7	A4 Background	3	0	0	0.172	0.125
25 °C	Z3	E7	Stripe	3	0	0	0.156	0.150
25 °C	Z3	E7	Trident	2	1	2	0.140	0.118

Table S3. Details of the QTLs detected in mapping experiments in both temperatures. The column °**C** indicates the temperature treatment. Effect indicates the effect size of the QTL peak. Chr indicates the chromosome. QTL C.I. Stand and End correspond to the boundaries of the confidence intervals, and positions for 2R and 3R continue from the last position on 2L and 3L, respectively. On the Pop Gen Outlier Genes column, "not analyzed" means that that QTL was in a region of low recombination and therefore was not analyzed for signatures of selection.

°C	Cross	Trait	Effect	Chr	QTL C.I. Start	QTL C.I. End	Pigmentation Genes	Pop Gen Outlier Genes
25	Z1E1	A4 back	0.25	X	8,781,757	8,884,188		
25	Z1E1	A4 back	0.15	2	0	1,777,569	*CG17650, CG43402*	*lea*
25	Z1E1	A4 back	0.10	3	0	2,671,607	*mwh, CG1887, klar, Ptp61F, Glut1, CG9134, bab2, Klp61F, bab1, CG7852*	*CG32483, CG3386, Lsp1gamma, Mkp, CG7028, ebd1, thoc7, RabX6, pyx, CG13405, rno, CG7049, Rev1, CG33229, p130CAS, Vdup1, CG43151, CG13875, NitFhit, CG13895, Kaz1-ORFB, Vti1, CG6845, Dic61B, CG43149, CG13876, CG3344, mthl8, Gyk, Mtch, CG42846, CG34140, CG3402, CG13894, mri,*

								CG16940, MED30, CG17129, CG12483, CG43150, CG13877, Pdk1
25	Z1E1	A4 back	0.15	3	40,938,289	42,148,205	*burs, e*	*Lrrk, CG7922, lbe*
25	Z2E1	A4 back	0.28	X	539,113	2,374,488	*Hr4*	not analyzed
25	Z2E1	A4 back	0.19	X	9,069,517	9,314,068	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E1	A4 back	0.13	3	529,924	2,034,129	*mwh, CG1887, klar, Ptp61F, Glut1, CG9134, bab2, Klp61F, bab1, CG7852*	*CG3402, CG32483, CG13895, CG3386, CG13894, Vti1, CG3344, MED30, CG17129, ebd1, Rev1, RabX6*
25	Z2E1	A4 back	0.35	3	41,460,800	41,614,751	*e*	
25	Z3E1	A4 back	0.11	3	15,004,647	31,196,314		not analyzed
25	Z3E1	A4 back	0.12	3	15,004,647	31,196,314		not analyzed
25	Z1E2	A4 back	0.25	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z1E2	A4 back	0.13	X	4,624,793	4,944,538	*CG42594, ovo*	
25	Z1E2	A4 back	0.16	X	9,180,719	9,484,036	*mgl*	*lz*
25	Z1E2	A4 back	0.12	3	12,525,223	12,613,606		
25	Z1E2	A4 back	0.10	3	19,510,104	32,491,294		not analyzed
25	Z2E2	A4 back	0.28	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z2E2	A4 back	0.16	X	9,094,352	9,424,829	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E2	A4 back	0.10	X	16,583,249	19,925,827	*vfl*	*CG12991, CG6873, CG32549, CCKLR-17D1, CG43759, CG12609, out, Diedel3, Hers, l(1)G0222, CG43996, CG8051, CG14190, CG43997, CG34326, CG34328*
25	Z2E2	A4 back	0.13	2	9,024,397	9,265,800		
25	Z2E2	A4 back	0.09	2	10,532,066	12,466,323	*crol*	*CG6734, CG6746, Mal-B1, CG14933, Oatp33Ea, CG4988, Vha100-5, Rh5*
25	Z2E2	A4 back	0.20	2	26,821,994	27,932,929	*pdm3*	not analyzed
25	Z2E2	A4 back	0.10	2	30,913,664	33,333,468	*CG42663, CG13330, CG30485*	*CG8778, CG17574, CG6220, CG8785, CG18368*
25	Z2E2	A4 back	0.12	2	41,554,081	42,396,510	*stl, Klp59D*	*Klp59C, CG42703, Gr59d, Gr59c*
25	Z2E2	A4 back	0.24	3	694,003	1,075,121	*Glut1, bab1*	*CG13895, CG13894*
25	Z2E2	A4 back	0.07	3	28,328,417	31,938,446	*hth, dsx, Mkk4, MBD-like*	not analyzed
25	Z2E2	A4 back	0.06	3	33,714,873	35,333,402	*E5, jvl*	*CG9813, Sdr, CG8870, cv-c, CG14861, CG8461, HtrA2, CG34273, PdE1, mRpL11, foxo*
25	Z2E2	A4 back	0.30	3	41,585,920	41,937,853	*e*	*CG7922, lbe*
25	Z2E2	A4 back	0.09	3	43,450,838	48,391,773	*nAcRalpha-96Aa, Mpk2, CG4815, TwdlC, CG6420, lobo*	*Tl, CG31097, pnt, CG31098, CG33337, TwdlQ, CG31104, CG31300, beat-VII, CG13658, beat-IV, CG10182, Lerp, CG31102, CG11893, scrib*
25	Z3E2	A4 back	0.14	X	0	3,001,335	*y, Hr4*	not analyzed

25	Z3E2	A4 back	0.12	X	9,766,419	9,940,595		
25	Z3E2	A4 back	0.08	X	10,319,425	11,102,487		*X11Lbeta, CG12637*
25	Z1E3	A4 back	0.18	X	0	2,313,405	*y, Hr4*	not analyzed
25	Z1E3	A4 back	0.11	X	7,057,882	8,237,116		*CG43287, CG15478, CHES-1-like*
25	Z1E3	A4 back	0.14	2	912,801	2,061,929	*CG17650, CG15362, CG43402, CG7337*	*lea*
25	Z1E3	A4 back	0.10	3	39,691,445	39,736,727		
25	Z1E3	A4 back	0.12	3	41,124,788	42,157,917	*burs, e*	*CG7922, lbe*
25	Z2E3	A4 back	0.41	X	0	2,227,106	*y, Hr4*	not analyzed
25	Z2E3	A4 back	0.17	X	6,983,307	7,632,850		*CG43287, CG15478, CHES-1-like*
25	Z2E3	A4 back	0.14	3	9,291,460	10,329,440	*CG32052, dpr6, Nc, dpr10*	*can, Klp67A, Hsp67Bb, Hsp22, CG4447, CG10809, CG32053, Or67d, CG32040, UGP, CG12362, Hsp67Bc, fry, Fdxh, CG32054, CG33696, CG14160, RasGAP1, CG4452*
25	Z2E3	A4 back	0.13	3	33,610,626	34,225,387		*CG9813, CG8870*
25	Z2E3	A4 back	0.28	3	41,574,540	41,924,132	*e*	*CG7922, lbe*
25	Z1E4	A4 back	0.16	X	8,950,905	9,301,924	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E4	A4 back	0.12	2	5,374,526	6,801,906	*CG8965, H15*	*ppk7, Lam, Hel25E, CG9171, CG9505, Oscillin, CG42368, CG14015, CG9498, slam*
25	Z1E4	A4 back	0.11	3	31,760,157	36,083,543	*E5, CG43336, jvl*	*CG11668, ry, snk, GstD5, CG7518, kar, CG18764, Cyp304a1, CG10038, CG43063, hug, CG43630, HtrA2, CG6188, GstD11, GstD9, CG8461, CG12594, mbo, GstD10, CG10035, Sdr, Spc25, Jupiter, CG14712, CG6959, CG34402, CG6813, l(3)neo38, CG6118, Octbeta3R, CG44037, PdE1, CG43062, CG10013, CG11656, CG10096, CG6225, Cyp313a4, grsm, CG14711, CG5404, GstD4, mRpL11, dpr17, CG8870, Octbeta2R, CG31446, beat-Va, CG33098, CG6808, CG34273, CG14395, GstD2, pxb, GstD6, CG17738, sim, CG6923, CG9813, Atx2, GstD7, Sfp87B, CG14861, CG10041, CG4702, lig3, CG14720, CG11670, CG5399, CG14384, CG42542, cv-c, GstD8, CG10097, svp, CG6752, CG14710, foxo, GstD1, Tim17a1, Lk6, CG8031*
25	Z1E4	A4 back	0.22	3	41,086,790	41,603,652	*e*	
25	Z2E4	A4 back	0.12	3	40,773,709	42,720,452	*burs, e*	*Lrrk, MtnD, Stat92E, CG7922, lbe*
25	Z1E5	A4 back	0.21	X	0	2,341,642	*y, Hr4*	not analyzed
25	Z1E5	A4 back	0.22	X	9,015,849	9,228,671	*t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E5	A4 back	0.14	X	8,903,940	9,354,420	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z3E5	A4 back	0.15	X	0	2,797,794	*y, Hr4*	not analyzed
25	Z1E6	A4 back	0.18	X	0	2,313,405	*y, Hr4*	not analyzed

25	Z1E6	A4 back	0.14	X	8,993,677	9,133,433	*t, Gr8a*	*CG12121, Gr8a, Ir8a*
25	Z1E6	A4 back	0.10	X	9,772,884	10,807,237		*X11Lbeta, CG12637*
25	Z2E6	A4 back	0.28	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z2E6	A4 back	0.22	X	9,043,226	9,202,381	*t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z3E6	A4 back	0.09	2	20,950,836	22,261,512	*CG31600, CG31702*	not analyzed
25	Z3E6	A4 back	0.06	2	25,381,793	27,329,638	*jing, CG1942, pdm3*	not analyzed
25	Z1E7	A4 back	0.21	X	0	2,332,069	*y, Hr4*	not analyzed
25	Z1E7	A4 back	0.22	X	8,888,363	9,027,375		
25	Z1E7	A4 back	0.10	2	10,152,895	13,057,341	*crol*	*CG9934, CG6734, CG6746, ACXE, Mal-B1, kek1, A16, CG14933, Oatp33Ea, CG4988, ACXC, Vha100-5, Tor, CG16800, Vha68-2, Rh5*
25	Z1E7	A4 back	0.10	3	40,307,800	44,027,872	*burs, Efa6, lmd, Sar1, e, loco*	*CG31213, CG4783, CG7922, loco, CG5023, beat-IV, Stat92E, CG10182, CG34139, Hs6st, pnt, CG6972, CG13842, Lrrk, CG42686, CG4367, lbe, CG42668, CG33337, MtnD, CG4362*
25	Z3E7	A4 back	0.11	X	0	3,159,367	*y, Hr4, dnc*	not analyzed
25	Z3E7	A4 back	0.10	X	3,251,402	4,903,588	*CG42594, bi*	
25	Z3E7	A4 back	0.17	X	9,109,204	9,396,397	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E1	Stripe	0.26	X	8,757,656	8,853,402		
25	Z1E1	Stripe	0.12	2	974,316	2,549,861	*CG15362, gho, CG7337, CG43402, CG17650*	*lea*
25	Z2E1	Stripe	0.28	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z2E1	Stripe	0.14	X	9,103,130	9,277,725	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E1	Stripe	0.17	3	694,003	1,462,791	*mwh, Ptp61F, Glut1, CG9134, bab2, Klp61F, bab1*	*CG13895, CG13894*
25	Z2E1	Stripe	0.11	3	4,027,121	7,155,756	*ple, Lkr, sinu, Sucb, CG10625, vvl*	*Lcp65Ag1, CG32241, Cpr65Ay, CG7465, CG4669, shep, Lcp65Aa, GluRIA, Lcp65Ab1, Cpr65Ax1, CG11349, CG13722, l(3)mbn, Lcp65Ab2, Cpr65Ax2, Lcp65Ag2, Acp65Aa, CG10576, Cpr65Az, Lcp65Af, CG13297, Lcp65Ae, CG32249, eIF4E-4, Ppat-Dpck, CG11350, Txl, Lcp65Ac, CG10226, CG32248, Cpr65Au, Mdr65, Lcp65Ag3*
25	Z2E1	Stripe	0.28	3	40,833,583	41,284,826		*Lrrk, MtnD, Stat92E*
25	Z1E2	Stripe	0.29	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z1E2	Stripe	0.11	X	4,409,208	5,423,661	*CG42594, ovo*	*CG42749, rg*
25	Z1E2	Stripe	0.10	X	6,135,717	6,740,827	*CG42340*	
25	Z1E2	Stripe	0.17	X	9,033,746	9,317,765	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E2	Stripe	0.13	2	30,141,290	31,119,500	*en*	*CG30031, CG30025, alphaTry, gammaTry, thetaTry, etaTry, betaTry, epsilonTry*

25	Z1E2	Stripe	0.11	3	441,915	2,291,629	*mwh, CG1887, klar, Ptp61F, Glut1, CG9134, bab2, Klp61F, bab1, CG7852*	*CG3402, CG32483, CG13895, CG3386, CG13894, Vti1, CG3344, MED30, CG17129, ebd1, Rev1, RabX6*
25	Z2E2	Stripe	0.35	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z2E2	Stripe	0.14	X	8,967,419	9,409,358	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E2	Stripe	0.11	2	8,217,086	10,207,534	*numb, CG33298*	*CG17834, CG32984, CG18088, Sema-1a, Apoltp*
25	Z2E2	Stripe	0.21	3	1,075,122	1,480,686	*mwh, Ptp61F, CG9134, bab2, Klp61F, bab1*	
25	Z2E2	Stripe	0.16	3	41,146,859	42,108,768	*e*	*CG7922, lbe*
25	Z3E2	Stripe	0.11	X	0	3,231,614	*y, Hr4, dnc*	not analyzed
25	Z3E2	Stripe	0.08	X	3,254,594	4,108,396		
25	Z3E2	Stripe	0.15	X	9,202,382	9,751,341	*mgl*	*CG32698*
25	Z1E3	Stripe	0.16	X	0	2,278,004	*y, Hr4*	not analyzed
25	Z1E3	Stripe	0.10	X	7,043,139	8,357,039		*CG43287, CG15478, CHES-1-like*
25	Z2E3	Stripe	0.33	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z2E3	Stripe	0.12	X	17,909,498	20,579,095	*vfl*	*CG33487, CG33498, obst-A, out, CG12679, Peritrophin-A, sw, CG43759, CG12609, Hers, Sdic4, hydra, CG34328, CG6873, CG32549, CCKLR-17D1, CG1835, Inx6, Diedel3, CG8051, Cyp6v1, CG1504, CG42577, Sdic3, CG14190*
25	Z1E4	Stripe	0.12	X	2,685,542	3,728,630	*dnc*	
25	Z1E4	Stripe	0.14	X	8,899,896	9,346,475	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E5	Stripe	0.24	X	9,103,130	9,277,725	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E5	Stripe	0.04	X	9,491,935	10,538,398		*CG32698*
25	Z2E5	Stripe	0.15	X	8,945,497	9,314,068	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z3E5	Stripe	0.12	X	9,584,933	10,876,962		*CG32698, X11Lbeta, CG12637*
25	Z1E6	Stripe	0.22	X	0	2,244,422	*y, Hr4*	not analyzed
25	Z1E6	Stripe	0.14	X	8,919,439	9,331,159	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E6	Stripe	0.27	X	0	2,354,025	*y, Hr4*	not analyzed
25	Z2E6	Stripe	0.19	X	8,945,497	9,207,359	*t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E7	Stripe	0.31	X	0	2,234,085	*y, Hr4*	not analyzed
25	Z1E7	Stripe	0.17	X	3,880,318	4,448,821	*bi*	
25	Z1E7	Stripe	0.17	X	8,804,662	8,972,939		
25	Z3E7	Stripe	0.16	X	0	2,584,808	*y, Hr4*	not analyzed
25	Z3E7	Stripe	0.15	X	3,565,208	4,833,995	*CG42594, bi*	

25	Z3E7	Stripe	0.15	X	9,027,376	9,442,114	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E1	Trident	0.19	X	8,999,627	9,284,598	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E1	Trident	0.11	2	40,551,327	42,215,000	*Klp59D, stl, PpD5, Fili*	*Klp59C, CG42703, Gr59d, Gr59c*
25	Z1E1	Trident	0.41	3	41,284,827	41,525,811		
25	Z2E1	Trident	0.25	X	0	2,227,106	*y, Hr4*	not analyzed
25	Z2E1	Trident	0.12	X	9,109,204	9,291,008	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E1	Trident	0.39	3	41,469,173	41,633,046	*e*	
25	Z3E1	Trident	0.10	3	37,820,325	38,093,297	*osa*	*osa, Pxt*
25	Z3E1	Trident	0.04	3	38,509,268	40,053,459	*CG15803*	
25	Z3E1	Trident	0.04	3	40,067,824	40,426,903	*Pk92B*	*CG31213, CG4783, Naam, CG34286, Hs6st*
25	Z3E1	Trident	0.13	3	40,705,432	41,805,035	*e*	*Lrrk, MtnD, Stat92E*
25	Z1E2	Trident	0.22	X	9,094,352	9,291,008	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E2	Trident	0.15	2	42,477,149	42,726,895		
25	Z1E2	Trident	0.14	3	4,755,865	5,610,217	*Sucb, CG10625, Lkr, sinu*	*shep, CG4669*
25	Z1E2	Trident	0.27	3	41,625,850	42,053,415		*CG7922, lbe*
25	Z2E2	Trident	0.12	X	704,754	3,251,401	*Hr4, dnc*	not analyzed
25	Z2E2	Trident	0.10	X	7,607,800	7,827,870		
25	Z2E2	Trident	0.19	X	9,113,708	9,367,505	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E2	Trident	0.12	2	8,774,319	8,956,837		
25	Z2E2	Trident	0.12	2	10,528,186	13,640,814	*crol*	*kek1, Oatp33Ea, CG33645, CG33640, Tor, Vha68-2, CG33641, CG16852, A16, CG14933, CG4988, Vha100-5, CG15638, CG33642, CG16800, CG6734, Mal-B1, ACXC, CG9934, CG6746, ACXE, B4, Rh5*
25	Z2E2	Trident	0.16	2	24,882,155	27,584,059	*jing, CG1942, pdm3*	not analyzed
25	Z2E2	Trident	0.09	2	30,577,618	33,713,959	*CG42663, CG13330, CG30485*	*blos1, CG33468, RpI1, CG17574, CG8778, CG6220, Su(var)2-HP2, CG33469, L, CG12868, tra2, CG8785, CG18368*
25	Z2E2	Trident	0.11	2	39,611,716	41,346,736	*otp, Pu, PpD5, Fili*	*CG30389, CG4266*
25	Z2E2	Trident	0.18	3	0	990,156	*Glut1, klar*	not analyzed
25	Z2E2	Trident	0.11	3	9,049,855	11,061,707	*dpr10, GluRIB, dpr6, Nc, CG32052, NijA*	*can, Klp67A, Hsp67Bb, klu, Hsp22, CG4447, CG10809, CG32053, Or67d, CG32040, UGP, CG12362, Hsp67Bc, fry, CG32079, Fdxh, CG32054, CG33696, CG14160, Fad2, RasGAP1, CG4452*
25	Z2E2	Trident	0.13	3	33,619,485	35,341,064	*E5, jvl*	*CG9813, Sdr, CG8870, cv-c, CG14861, CG8461, HtrA2, CG34273, PdE1, mRpL11, foxo*
25	Z2E2	Trident	0.31	3	41,582,176	41,931,294	*e*	*CG7922, lbe*
25	Z3E2	Trident	0.13	X	0	3,246,133	*y, Hr4, dnc*	not analyzed

25	Z3E2	Trident	0.09	X	9,291,009	12,038,929	*mgl, ATP7*	*CG2750, CG32698, CG15741, Cyp28c1, X11Lbeta, CG12637, CG15740, Ten-a*
25	Z3E2	Trident	0.15	X	9,291,009	12,038,929	*mgl, ATP7*	*CG2750, CG32698, CG15741, Cyp28c1, X11Lbeta, CG12637, CG15740, Ten-a*
25	Z3E2	Trident	0.10	3	16,451,860	32,246,885		not analyzed
25	Z3E2	Trident	0.10	3	33,458,602	39,708,929	*Abd-B, CG43336, E5, abd-A, CG15803, osa, jvl*	*CG8870, alpha-Man-IIb, CG4546, CG14861, CG4520, srp, Hel89B, CREG, CG5225, CG31446, Pxt, Sdr, CG34273, beat-IIa, CG5399, Pxd, CG31183, CG42542, pxb, CG6118, cv-c, lute, CG6752, CG4576, HtrA2, PdE1, cv-d, CG43196, foxo, CG9813, CG12784, osa, Hmt-1, sds22, CG31419, rec, Atx2, CG5404, CG8461, Brf, mRpL11, Sur-8*
25	Z3E2	Trident	0.11	3	33,458,602	39,708,929	*Abd-B, CG43336, E5, abd-A, CG15803, osa, jvl*	*CG8870, alpha-Man-IIb, CG4546, CG14861, CG4520, srp, Hel89B, CREG, CG5225, CG31446, Pxt, Sdr, CG34273, beat-IIa, CG5399, Pxd, CG31183, CG42542, pxb, CG6118, cv-c, lute, CG6752, CG4576, HtrA2, PdE1, cv-d, CG43196, foxo, CG9813, CG12784, osa, Hmt-1, sds22, CG31419, rec, Atx2, CG5404, CG8461, Brf, mRpL11, Sur-8*
25	Z3E2	Trident	0.07	3	40,658,762	40,705,431		*CG42668, CG4367, CG4362*
25	Z3E2	Trident	0.05	3	40,833,583	41,445,463		*Lrrk, MtnD, Stat92E*
25	Z3E2	Trident	0.09	3	40,833,583	41,445,463		*Lrrk, MtnD, Stat92E*
25	Z3E2	Trident	0.05	3	41,445,466	41,556,938		
25	Z3E2	Trident	0.04	3	41,556,941	42,265,267	*burs, e*	*CG7922, lbe*
25	Z3E2	Trident	0.10	3	47,871,919	51,744,389	*CG14506, kay, CG15550, CecC, CG43448, CG4815, CG1340*	*Jon99Ci, CG14061, Jon99Cii, Cog7, trp, CG15522, Jon99Ciii, CG9997, beat-VI, CG11873, capa, CG11340, CG15555, CG31202, CG10000, CG1894, CG34295*
25	Z1E3	Trident	0.12	2	0	2,274,740	*CG15362, gho, CG7337, CG43402, CG17650*	*lea*
25	Z1E3	Trident	0.09	3	39,714,854	39,803,049		
25	Z1E3	Trident	0.23	3	41,614,752	42,138,272		*CG7922, lbe*
25	Z2E3	Trident	0.31	X	0	2,278,004	*y, Hr4*	not analyzed
25	Z2E3	Trident	0.11	X	6,697,815	8,268,842		*CG43287, CG15478, CHES-1-like*
25	Z2E3	Trident	0.11	3	1,583,911	3,756,863	*dar1, CG1887*	*CG12017*
25	Z2E3	Trident	0.15	3	34,445,578	34,639,471		
25	Z2E3	Trident	0.32	3	41,574,540	41,924,132	*e*	*CG7922, lbe*
25	Z2E3	Trident	0.08	3	46,212,783	48,241,893	*TwdlC, CG6420*	*Tl, TwdlQ, beat-VII, Lerp, scrib*
25	Z3E3	Trident	0.13	X	0	3,085,825	*y, Hr4, dnc*	not analyzed
25	Z3E3	Trident	0.10	3	39,317,300	39,441,402		
25	Z3E3	Trident	0.11	3	42,398,434	43,515,960	*lmd, loco, Efa6, Sar1*	*CG13842, CG42686, CG6972, loco*
25	Z3E3	Trident	0.06	3	43,783,598	46,343,503	*nAcRalpha-96Aa, Mpk2, lobo*	*CG31097, CG31098, CG33337, CG31104, CG13658, beat-IV, CG10182, CG31102,*

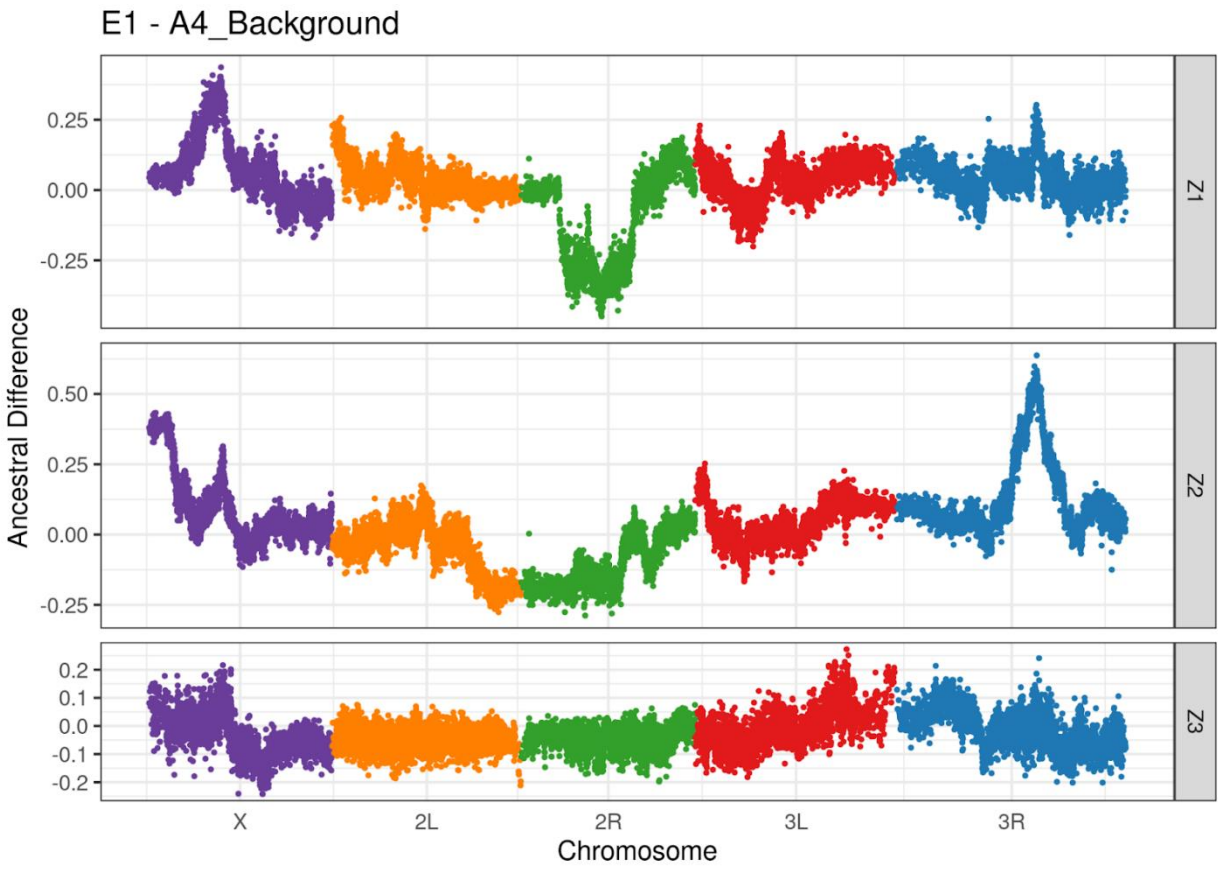
								CG11893, CG31300
25	Z1E4	Trident	0.13	3	24,543,556	28,202,986		not analyzed
25	Z1E4	Trident	0.05	3	33,815,390	34,516,931	*E5*	*foxo*
25	Z1E4	Trident	0.19	3	40,860,812	41,205,158		*Lrrk, MtnD, Stat92E*
25	Z1E4	Trident	0.17	3	41,205,161	41,536,427		
25	Z2E4	Trident	0.17	3	41,484,713	42,473,754	*burs, e*	*CG7922, lbe*
25	Z3E4	Trident	0.08	3	14,457,784	14,937,152		*mop, CG9384, CG17173*
25	Z3E4	Trident	0.07	3	14,939,884	16,267,201		*Best4, l(3)72Dh, SsRbeta, CG5151, CG33795, CG33689, CG33688, CG12272, CG33687, CG5157, Pgm, CG16838, sff, CG33259, CG32152, Toll-6, Pka-C3, Tollo*
25	Z3E4	Trident	0.07	3	16,753,965	17,128,032	*Rbp6*	*CG9669, TSG101, CG13024, CG9715, CG32161, Rbp6, Lmpt, Rh4, nudC*
25	Z3E4	Trident	0.06	3	17,678,351	30,598,201		not analyzed
25	Z3E4	Trident	0.12	3	17,678,351	30,598,201		not analyzed
25	Z3E4	Trident	0.21	3	17,678,351	30,598,201		not analyzed
25	Z3E4	Trident	0.13	3	39,330,425	39,461,873		
25	Z3E4	Trident	0.08	3	39,478,566	41,972,910	*e, Pk92B*	*CG42668, CG31213, Lrrk, MtnD, CG4783, CG5023, Naam, Stat92E, CG34286, CG34139, CG4367, CG7922, Hs6st, CG4362, lbe*
25	Z3E4	Trident	0.06	3	42,024,865	43,469,768	*burs, Efa6, lmd, Sar1, loco*	*CG13842, CG42686, CG6972, loco*
25	Z3E4	Trident	0.13	3	43,469,771	43,686,837		
25	Z1E5	Trident	0.29	X	9,043,226	9,196,729	*t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E5	Trident	0.07	X	9,491,935	10,617,378		*CG32698, X11Lbeta, CG12637*
25	Z1E5	Trident	0.13	2	35,887,484	36,109,611		*Cda9, CG11400, Amy-p, CG11395, CG15605, CG15611, Spn53F, Amy-d, Gbp*
25	Z1E5	Trident	0.12	2	39,661,087	40,431,788	*otp, Pu*	*CG30389, CG4266*
25	Z1E5	Trident	0.11	3	41,284,827	41,521,262		
25	Z1E5	Trident	0.08	3	41,683,254	42,265,267	*burs*	*CG7922, lbe*
25	Z2E5	Trident	0.18	X	8,987,090	9,277,725	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z3E5	Trident	0.11	X	0	3,410,979	*y, Hr4, dnc*	not analyzed
25	Z3E5	Trident	0.13	X	9,967,074	10,180,813		
25	Z3E5	Trident	0.10	3	15,915,000	32,349,700		not analyzed
25	Z3E5	Trident	0.14	3	15,915,000	32,349,700		not analyzed
25	Z3E5	Trident	0.06	3	41,438,173	41,594,336		
25	Z3E5	Trident	0.04	3	41,666,679	42,359,982	*burs*	*CG7922, lbe*
25	Z3E5	Trident	0.04	3	42,359,985	43,594,625	*lmd, loco, Efa6, Sar1*	*CG13842, CG42686, CG6972, loco*

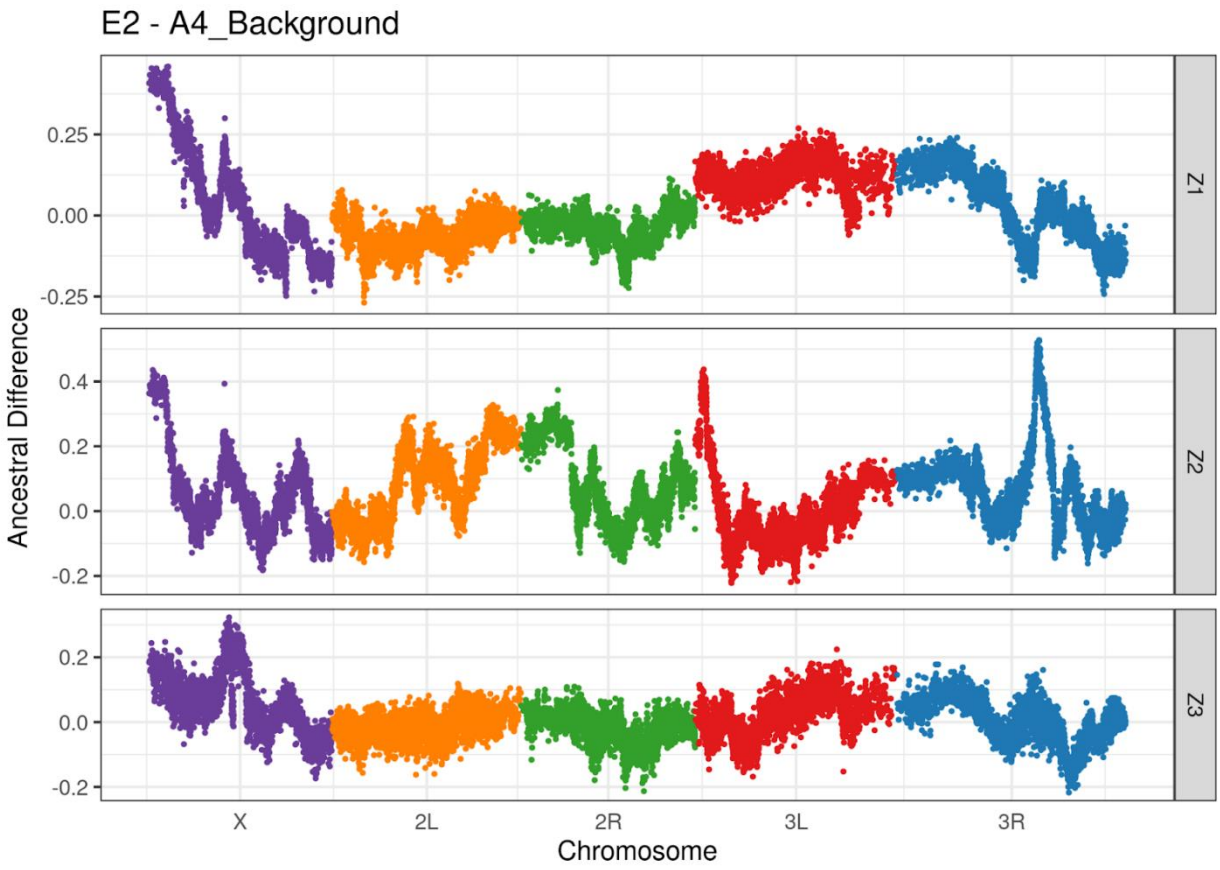
25	Z1E6	Trident	0.12	X	0	3,213,540	*y, Hr4, dnc*	not analyzed
25	Z1E6	Trident	0.11	X	4,802,907	5,765,347	*ovo*	*CG42749, rg*
25	Z1E6	Trident	0.15	X	8,909,003	9,270,976	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E6	Trident	0.20	X	0	2,323,687	*y, Hr4*	not analyzed
25	Z2E6	Trident	0.37	X	9,043,226	9,202,381	*t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z2E6	Trident	0.14	3	45,972,800	47,575,467	*TwdlC, CG6420*	*Tl, TwdlQ, beat-VII, Lerp, scrib*
25	Z3E6	Trident	0.21	X	9,196,730	9,375,114	*mgl*	
25	Z3E6	Trident	0.06	2	20,950,836	22,081,080	*CG31600, CG31702*	not analyzed
25	Z3E6	Trident	0.04	2	22,081,083	24,299,472		not analyzed
25	Z3E6	Trident	0.05	2	24,304,316	25,350,319		not analyzed
25	Z3E6	Trident	0.05	2	25,350,322	26,320,904	*jing*	not analyzed
25	Z3E6	Trident	0.14	3	40,773,709	42,119,462	*e*	*Lrrk, MtnD, Stat92E, CG7922, lbe*
25	Z1E7	Trident	0.15	X	8,972,940	9,381,975	*mgl, t, Gr8a*	*CG12121, Gr8a, Ir8a, lz*
25	Z1E7	Trident	0.11	X	19,261,637	22,198,452	*vfl*	*CG33487, CG33498, CG32820, Cyp6t1, CG12446, CG14618, CG14613, l(1)G0196, obst-A, DIP1, slgA, CG1718, CG14619, CG12576, Cp110, CG12679, CG17599, Peritrophin-A, sw, CG14614, tilB, Hers, Sdic4, hydra, CG11227, CG17600, S6kII, Ir20a, CG1835, CG32819, CG14615, Inx6, CG14621, CG17601, Cyp6v1, CG1504, CG32857, CG32500, CG14476, Cda4, CG33502, CG15450, waw, CG42577, CG17598, CG17450, shakB, bbx, Sdic3*
25	Z1E7	Trident	0.24	2	12,047,685	12,256,828		
25	Z1E7	Trident	0.10	3	6,743,111	10,396,278	*dpr10, TrpA1, Gug, GluRIB, RecQ4, dpr6, Nc, Doc2, vvl, CG32052, dally*	*can, mfr, Klp67A, Hsp67Bb, Hsp22, unc-13-4A, CG4447, CG10809, CG32053, Or67d, CG32040, UGP, TrpA1, CG12362, Hsp67Bc, fry, Fdxh, CG32054, Ect4, CG33696, CG14160, RasGAP1, CG4452*
25	Z1E7	Trident	0.20	3	41,319,048	42,017,920	*e*	*CG7922, lbe*
25	Z3E7	Trident	0.13	X	0	3,246,133	*y, Hr4, dnc*	not analyzed
25	Z3E7	Trident	0.14	X	7,778,868	8,543,491		
25	Z3E7	Trident	0.10	2	36,748,929	38,405,641	*sbb, edl, GEFmeso*	*CG18607, rib, Elk, CG18606, CG11906, CG5773, 5-HT1A, CG30114, CG10474, Hs3st-A, CG10476, CG34386, CG10910, FK506-bp2, CG33958*
25	Z3E7	Trident	0.11	3	8,982,154	10,927,857	*dpr10, GluRIB, dpr6, Nc, Doc2, CG32052, NijA*	*can, Klp67A, Hsp67Bb, Hsp22, CG4447, CG10809, CG32053, Or67d, CG32040, UGP, CG12362, Hsp67Bc, fry, Fdxh, CG32054, CG33696, CG14160, RasGAP1, CG4452*
25	Z3E7	Trident	0.11	3	40,430,381	42,897,349	*burs, Sar1, e*	*CG42668, Lrrk, MtnD, CG5023, Stat92E, CG34139, CG4367, CG7922, CG4362, lbe*
15	Z1E4	A4 back	0.27	X	8,956,934	9,133,433	*t, Gr8a*	*Gr8a, CG12121, Ir8a*
15	Z1E4	A4 back	0.12	3	16,862,363	31,938,446		not analyzed

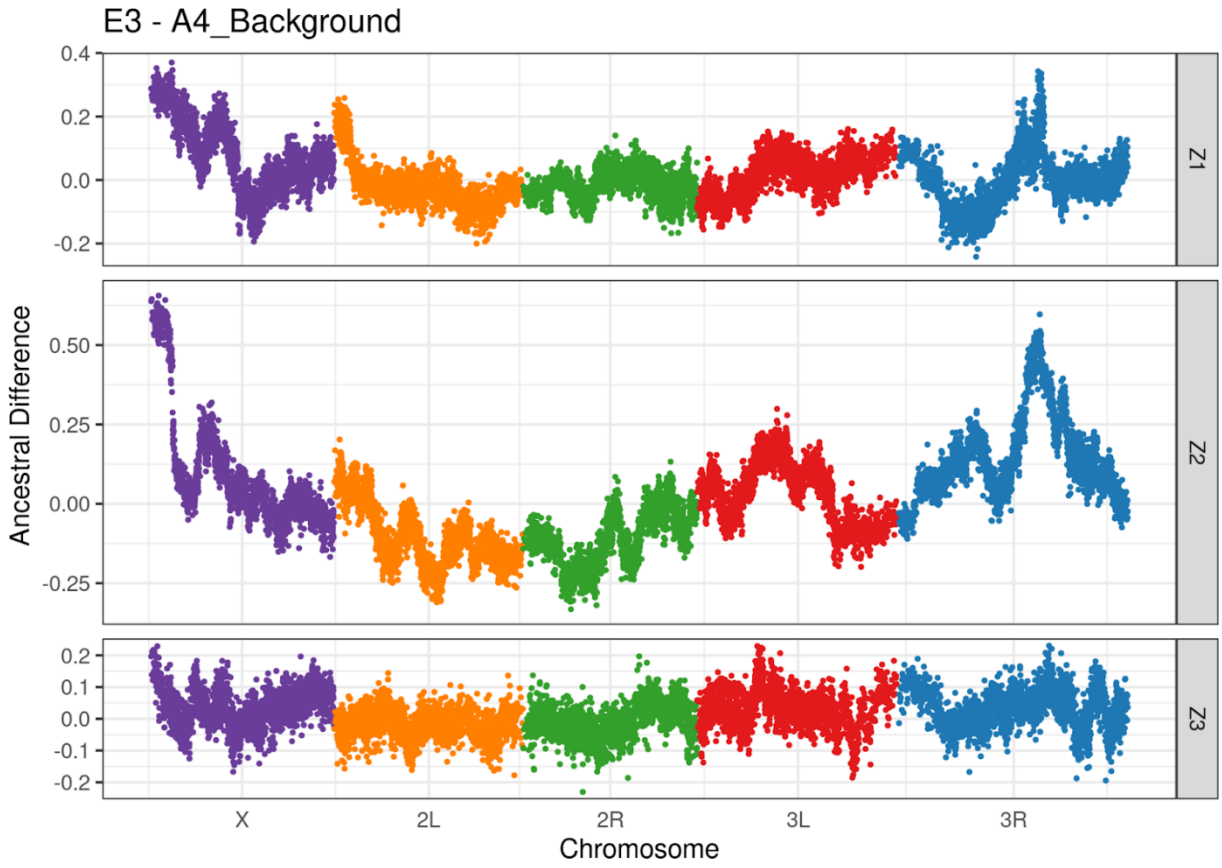
15	Z1E4	A4 back	0.37	3	41,291,992	41,540,188		
15	Z3E4	A4 back	0.14	X	0	2,752,084	*Hr4, y*	not analyzed
15	Z3E4	A4 back	0.17	X	2,188,454	2,823,419		*CG14050, phl*
15	Z3E4	A4 back	0.05	X	13,423,775	13,660,170		
15	Z3E4	A4 back	0.13	X	13,726,426	14,091,527		
15	Z3E4	A4 back	0.05	3	45,823,296	46,028,206		
15	Z3E4	A4 back	0.05	3	46,400,431	46,992,714		*scrib*
15	Z3E4	A4 back	0.04	3	46,996,027	48,177,291	*CG6420, TwdlC*	*TwdlQ, Tl, Lerp, beat-VII*
15	Z1E7	A4 back	0.34	X	0	2,234,085	*Hr4, y*	not analyzed
15	Z1E7	A4 back	0.27	X	8,872,688	8,950,904		
15	Z1E7	A4 back	0.11	2	40,697,109	42,306,895	*stl, Klp59D, PpD5, Fili*	*CG42703, Klp59C, Gr59d, Gr59c*
15	Z1E7	A4 back	0.16	3	40,814,889	41,830,695	*e*	*MtnD, lbe, CG7922, Lrrk, Stat92E*
15	Z3E7	A4 back	0.19	X	0	2,415,104	*Hr4, y*	not analyzed
15	Z3E7	A4 back	0.17	X	8,866,166	9,109,203		
15	Z1E4	Stripe	0.05	X	0	2,234,085	*Hr4, y*	not analyzed
15	Z1E4	Stripe	0.05	X	0	2,354,025	*Hr4, y*	not analyzed
15	Z1E4	Stripe	0.16	X	2,802,933	3,284,593	*dnc*	
15	Z3E4	Stripe	0.12	3	18,190,307	30,266,385		not analyzed
15	Z1E7	Stripe	0.40	X	0	2,234,085	*Hr4, y*	not analyzed
15	Z3E7	Stripe	0.16	X	0	3,091,113	*Hr4, dnc, y*	not analyzed
15	Z3E7	Stripe	0.11	X	3,284,594	4,301,912		
15	Z1E4	Tridendt	0.25	X	8,956,934	9,180,718	*t, Gr8a*	*Gr8a, CG12121, Ir8a*
15	Z1E4	Tridendt	0.18	3	9,147,755	9,883,377	*GluRIB, CG32052*	*fry, RasGAP1, Fdxh, CG4452, Hsp22, CG33696, Hsp67Bc, CG10809, UGP, Klp67A, CG4447, CG32040, Hsp67Bb*
15	Z1E4	Tridendt	0.16	3	19,429,112	29,490,441		not analyzed
15	Z1E4	Tridendt	0.40	3	40,860,812	41,230,656		*Lrrk, Stat92E, MtnD*
15	Z3E4	Tridendt	0.11	3	3,229,826	6,369,714	*CG10625, Sucb, sinu, scrt, dar1, Lkr*	*Ppat-Dpck, CG12017, CG32248, Lcp65Ae, Acp65Aa, Lcp65Af, CG4669, Cpr65Au, Cpr65Ax1, CG7465, Lcp65Ag3, CG13297, CG10576, Cpr65Ax2, shep, Txl, Lcp65Ag1, Lcp65Ac, Cpr65Ay, CG11349, CG32249, CG10226, Lcp65Aa, CG13722, Cpr65Az, l(3)mbn, Lcp65Ab1, CG11350, Mdr65, CG32241, Lcp65Ab2, Lcp65Ag2*
15	Z3E4	Tridendt	0.15	3	13,980,269	15,696,541		*mop, CG33259, Toll-6, Tollo, CG17173, CG9384, Best4*
15	Z3E4	Tridendt	0.15	3	17,563,182	20,374,596	*Mi-2, Eip74EF, gogo*	not analyzed
15	Z3E4	Tridendt	0.11	3	39,305,383	39,426,310		

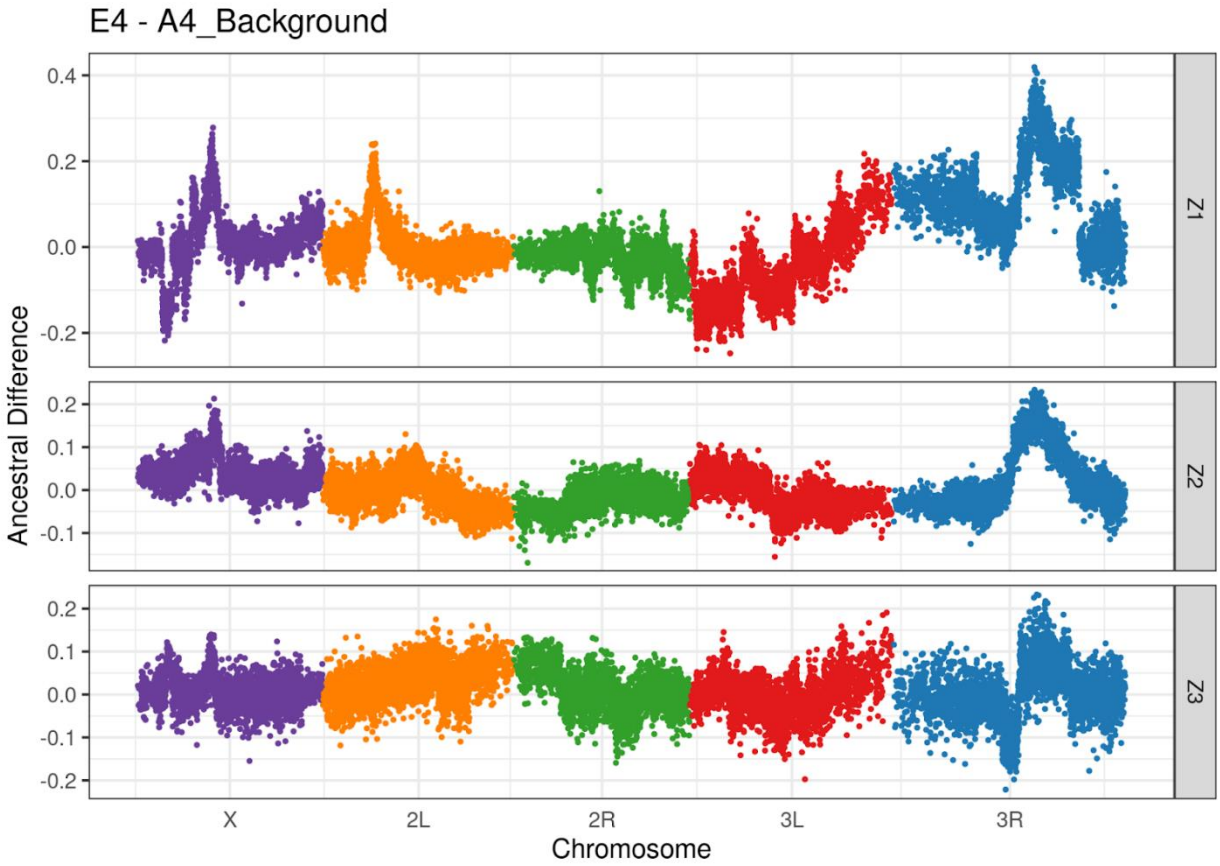
15	Z3E4	Tridendt	0.16	3	40,892,214	41,862,460	*e*	*MtnD, lbe, CG7922, Lrrk, Stat92E*
15	Z1E7	Tridendt	0.19	X	0	2,265,403	*Hr4, y*	not analyzed
15	Z1E7	Tridendt	0.33	X	8,926,127	9,007,996		
15	Z1E7	Tridendt	0.11	2	7,898,584	7,965,461		
15	Z1E7	Tridendt	0.09	2	11,752,348	14,025,027	*crol, p38b*	*CG6746, CG7968, Vha68-2, CG6734, CG33645, ACXE, CG33640, B4, Mal-B1, CG9934, CG16800, CenG1A, CG15638, Oatp33Ea, CG14933, Rh5, CG16852, CG33641, A16, CG33642, ACXC, Tor, kek1*
15	Z1E7	Tridendt	0.12	3	12,394,463	15,385,748		*Vps36, CG34429, mRpL20, CG17173, CG9384, CG17300, Liprin-beta, CG11267, stv, Acp70A, CG33259, Spt20, MICAL-like, CG14113, ste14, Best4, CG10710, mop, Toll-6, Tollo, bru-3, CG10089*
15	Z1E7	Tridendt	0.25	3	40,958,451	41,453,725		*Lrrk*
15	Z3E7	Tridendt	0.24	X	9,180,719	9,331,159	*mgl*	*lz*
15	Z3E7	Tridendt	0.10	3	39,166,096	39,267,567		
15	Z3E7	Tridendt	0.11	3	40,682,161	42,198,206	*burs, e*	*MtnD, lbe, CG7922, Lrrk, Stat92E*
15	Z3E7	Tridendt	0.10	3	47,958,436	51,485,456	*kay, CG15550, CecC, CG14506, CG43448, CG1340, CG4815*	*Jon99Ciii, beat-VI, capa, trp, Jon99Cii, CG10000, CG34295, CG14061, Jon99Ci, CG1894, CG9997, CG15522, Cog7, CG31202, CG11873*

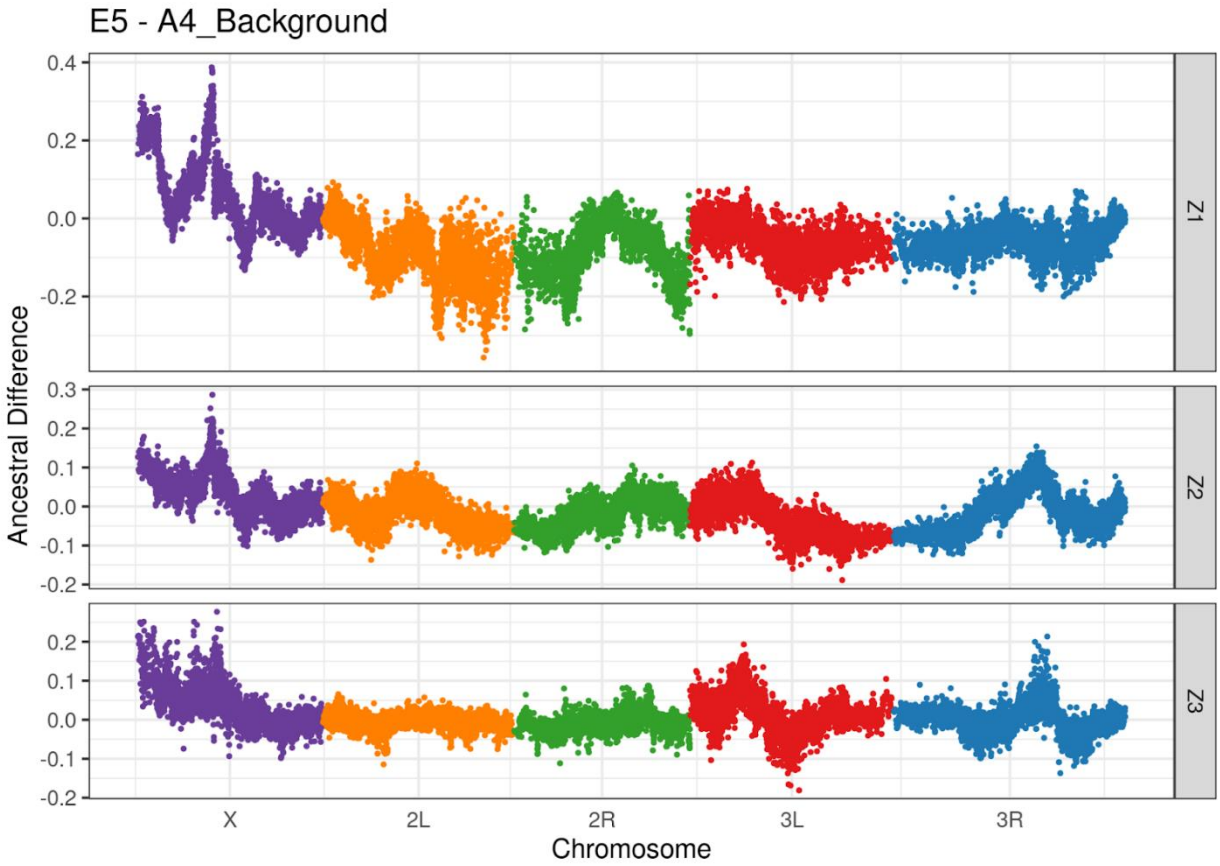
Figures S1, S2, and S3 are multipage figures showing genomewide ancestry difference distribution for all mapping experiments performed at 25 °C.

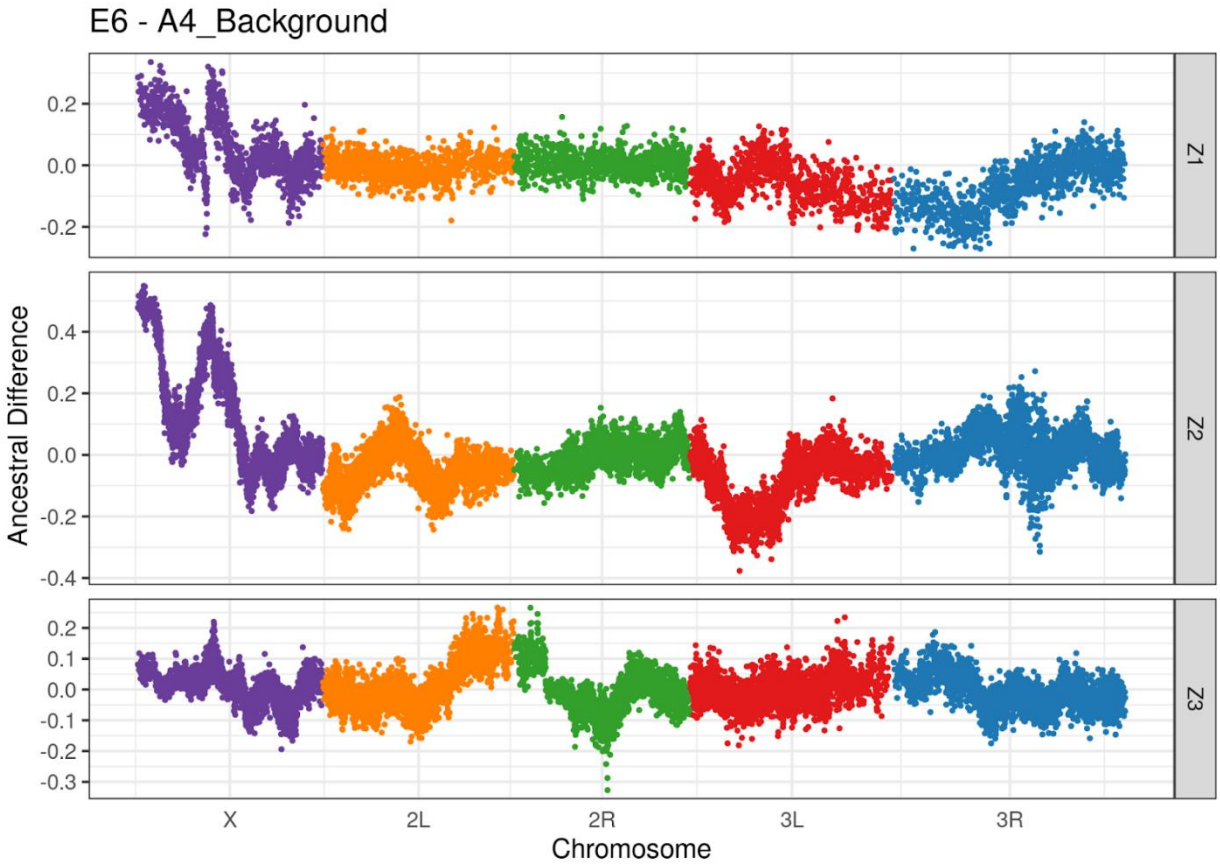

E1 - A4_Background
Ancestral Difference
0.25
0.00
-0.25
Z1
0.50
0.25
0.00
-0.25
Z2
0.2
0.1
0.0
-0.1
-0.2
Z3
X
2L
2R
3L
3R
Chromosome


E2 - A4_Background
Ancestral Difference
0.25
0.00
-0.25
0.4
0.2
0.0
-0.2
0.2
0.0
-0.2
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E3 - A4_Background
Ancestral Difference
0.4
0.2
0.0
-0.2
Z1
0.50
0.25
0.00
-0.25
Z2
0.2
0.1
0.0
-0.1
-0.2
Z3
X
2L
2R
3L
3R
Chromosome


E4 - A4_Background
Ancestral Difference
0.4
0.2
0.0
-0.2
Z1
0.2
0.1
0.0
-0.1
Z2
0.2
0.1
0.0
-0.1
-0.2
Z3
X
2L
2R
3L
3R
Chromosome


E5 - A4_Background
Ancestral Difference
0.4
0.2
0.0
-0.2
0.3
0.2
0.1
0.0
-0.1
-0.2
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E6 - A4_Background
Ancestral Difference
0.2
0.0
-0.2
0.4
-0.4
0.1
-0.1
-0.3
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome

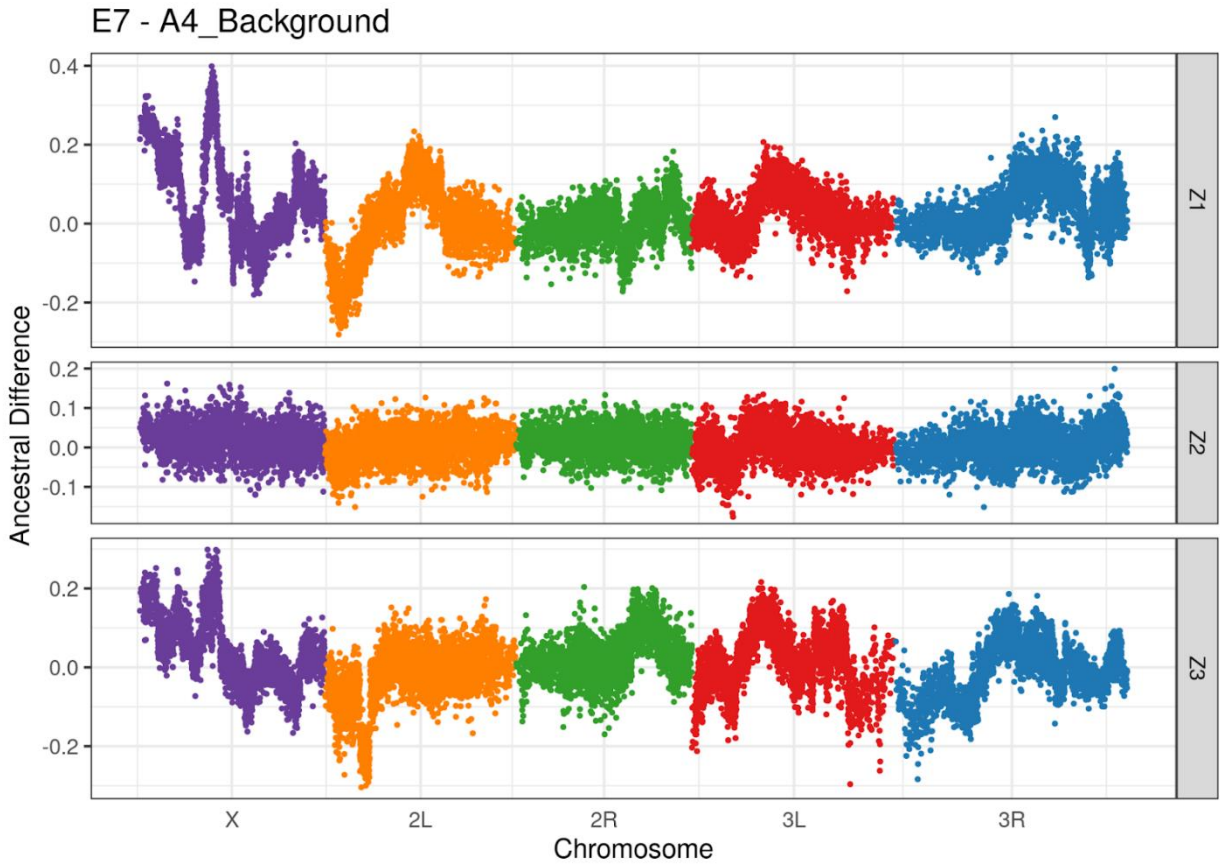
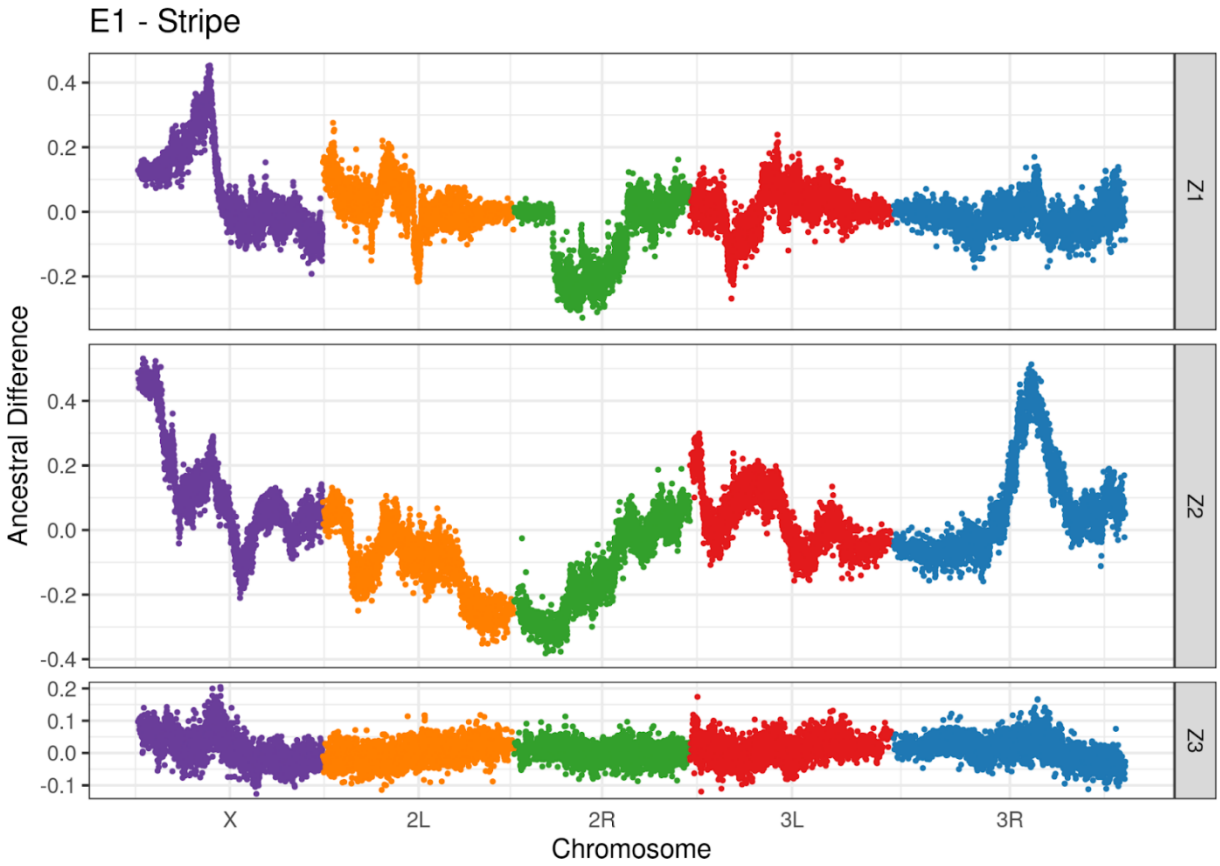
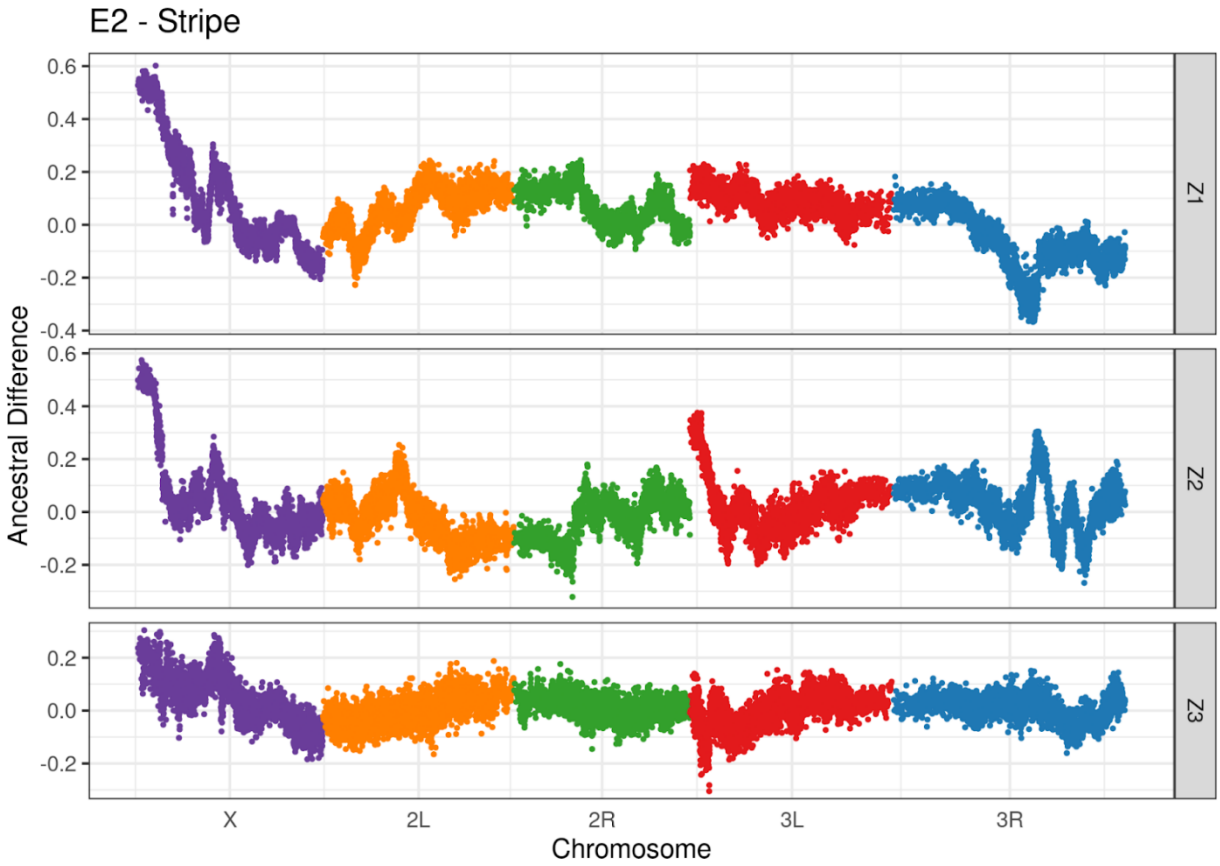
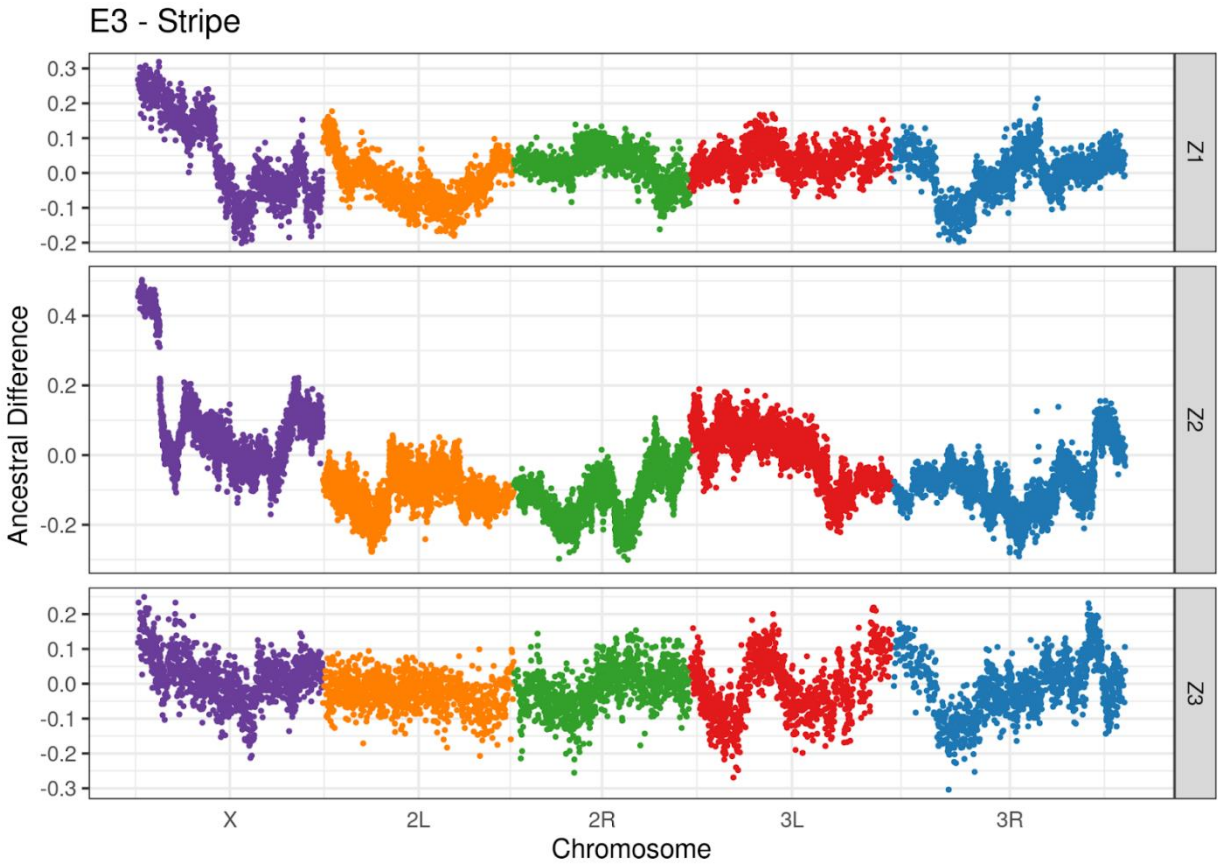
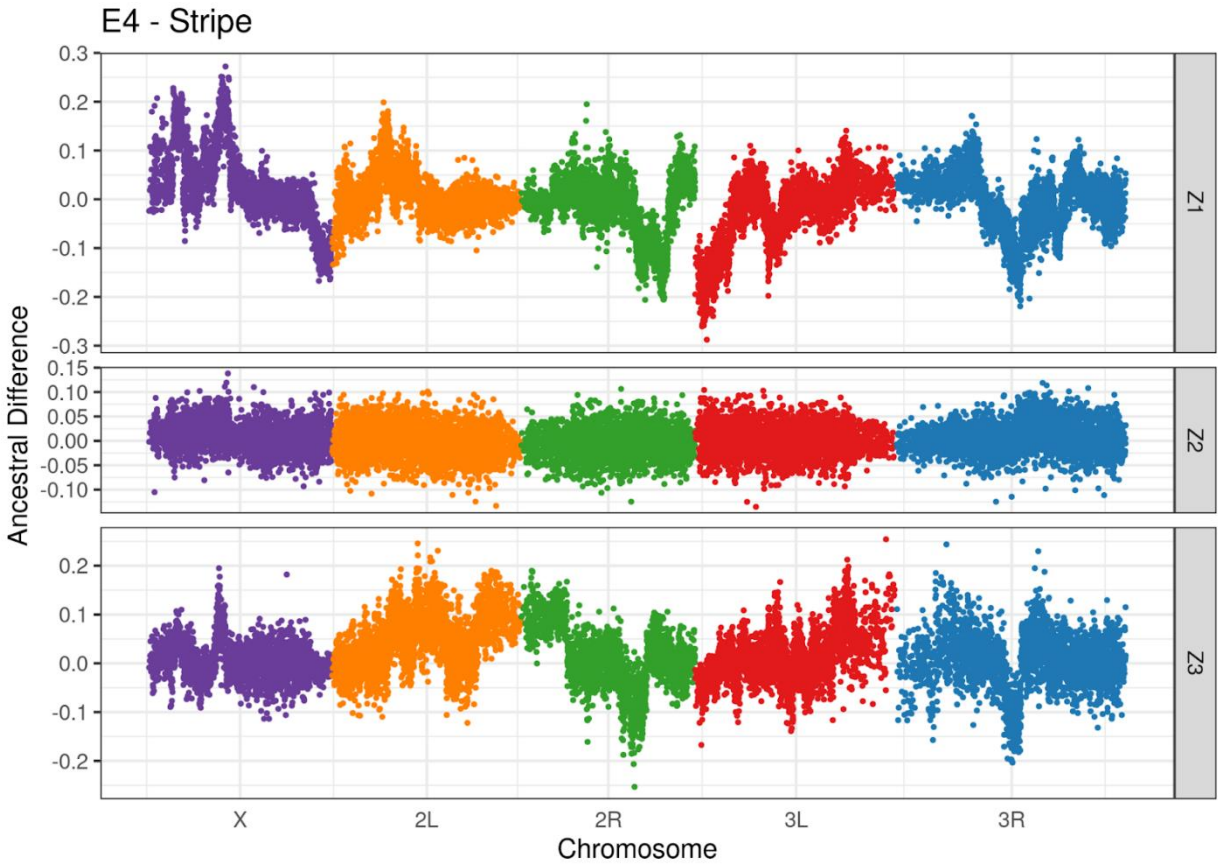


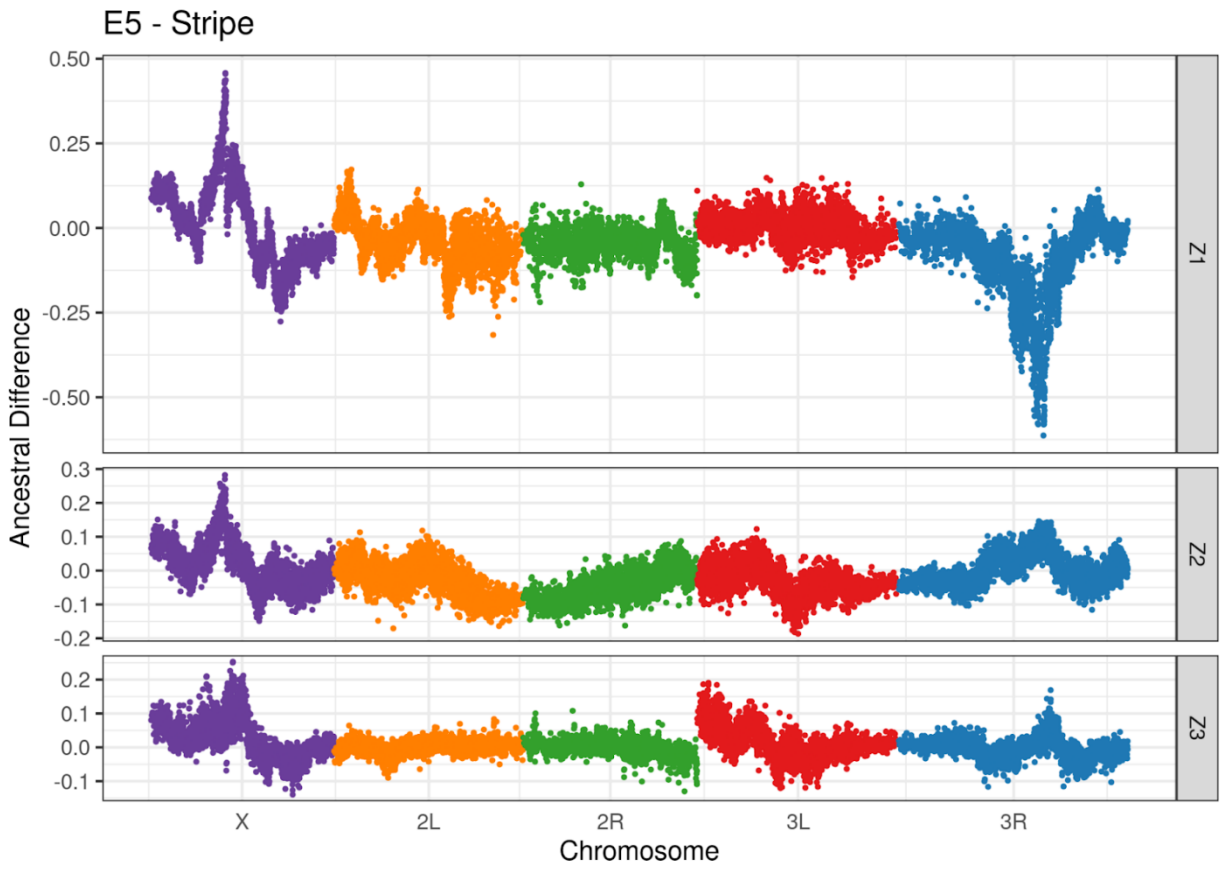
Figure S1. Genomewide ancestry difference distribution for A4 Background, different colors representing different chromosome arms. Each page shows three crosses, one Ethiopian by three Zambian parental strains. Ethiopian strain is shown on the top of the plot, and Zambian strains are shown on the right y-axis.

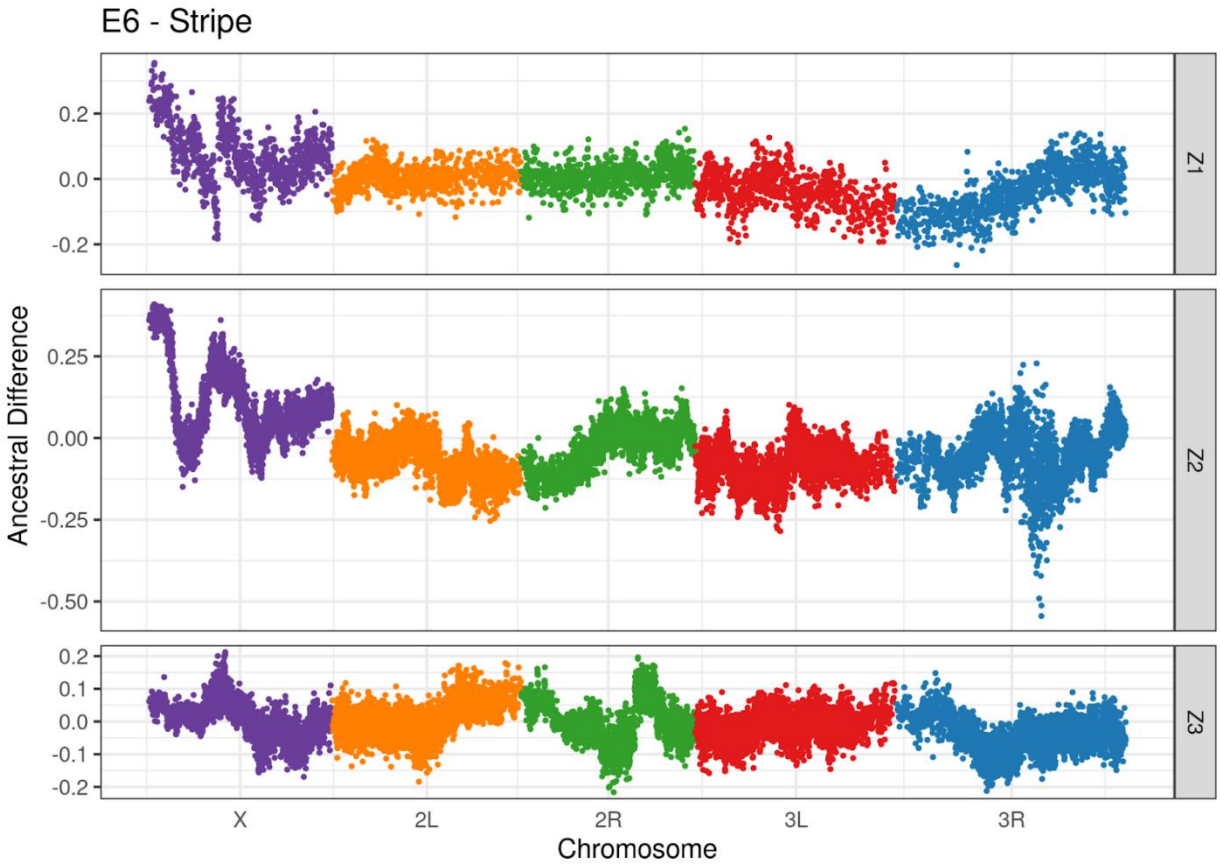

E1 - Stripe
Ancestral Difference
0.4
0.2
0.0
-0.2
0.4
0.2
0.0
-0.2
-0.4
0.2
0.1
0.0
-0.1
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E2 - Stripe
Ancestral Difference
0.6
0.4
0.2
0.0
-0.2
-0.4
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E3 - Stripe
Ancestral Difference
0.3
0.2
0.1
0.0
-0.1
-0.2
0.4
0.2
0.0
-0.2
0.2
0.1
0.0
-0.1
-0.2
-0.3
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E4 - Stripe
Ancestral Difference
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Z1
0.15
0.10
0.05
0.00
-0.05
-0.10
Z2
0.2
0.1
0.0
-0.1
-0.2
Z3
X
2L
2R
3L
3R
Chromosome


E5 - Stripe
Ancestral Difference
0.50
0.25
0.00
-0.25
-0.50
Z1
0.3
0.2
0.1
0.0
-0.1
-0.2
Z2
0.2
0.1
0.0
-0.1
Z3
X
2L
2R
3L
3R
Chromosome


E6 - Stripe
Ancestral Difference
0.2
0.0
-0.2
0.25
0.00
-0.25
-0.50
0.2
0.1
0.0
-0.1
-0.2
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome

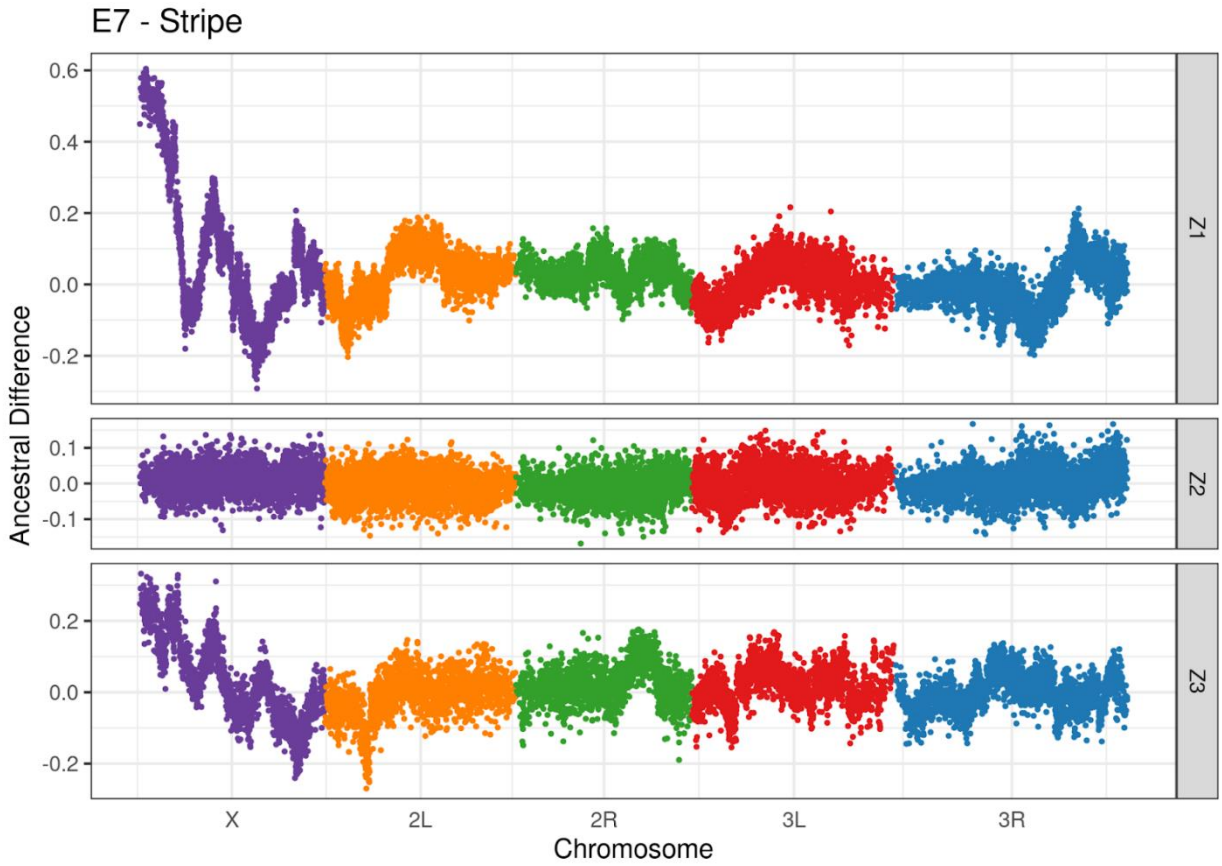
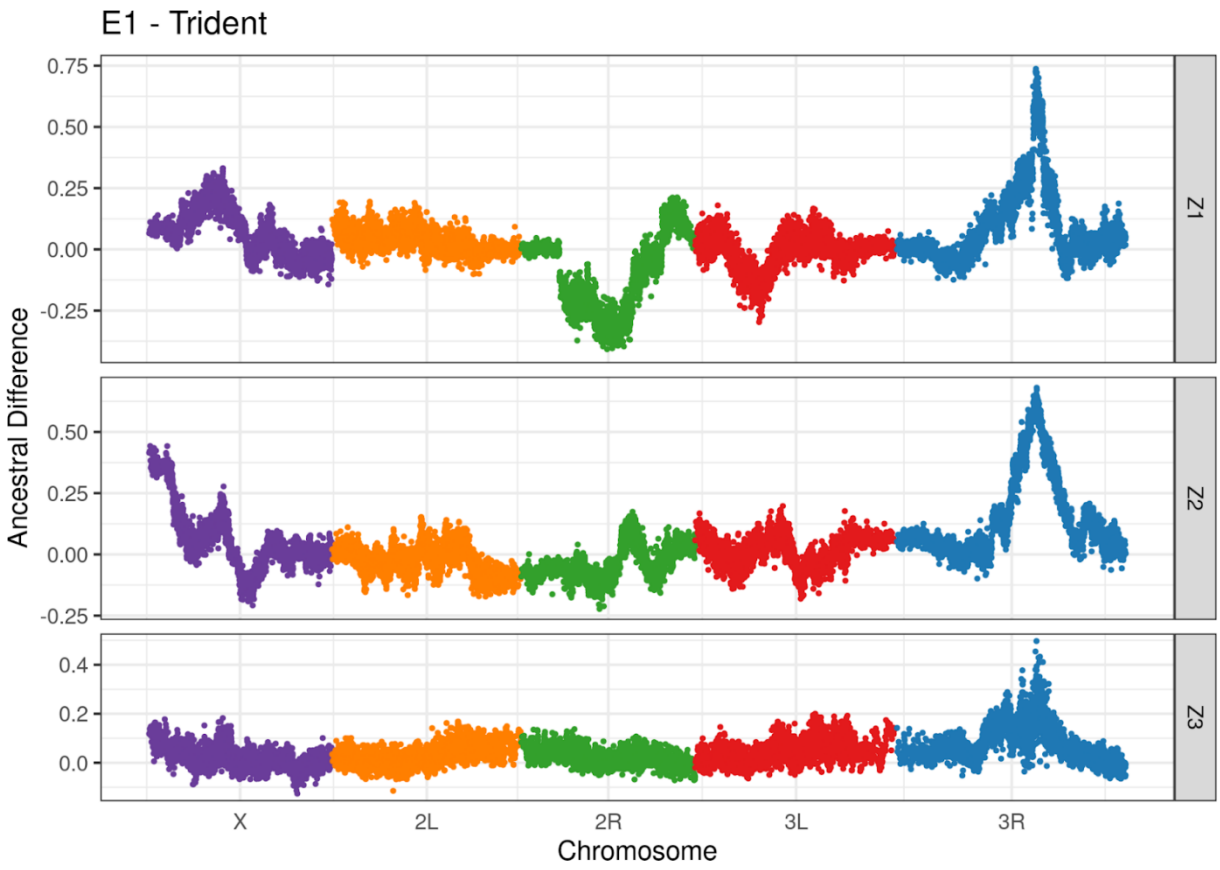
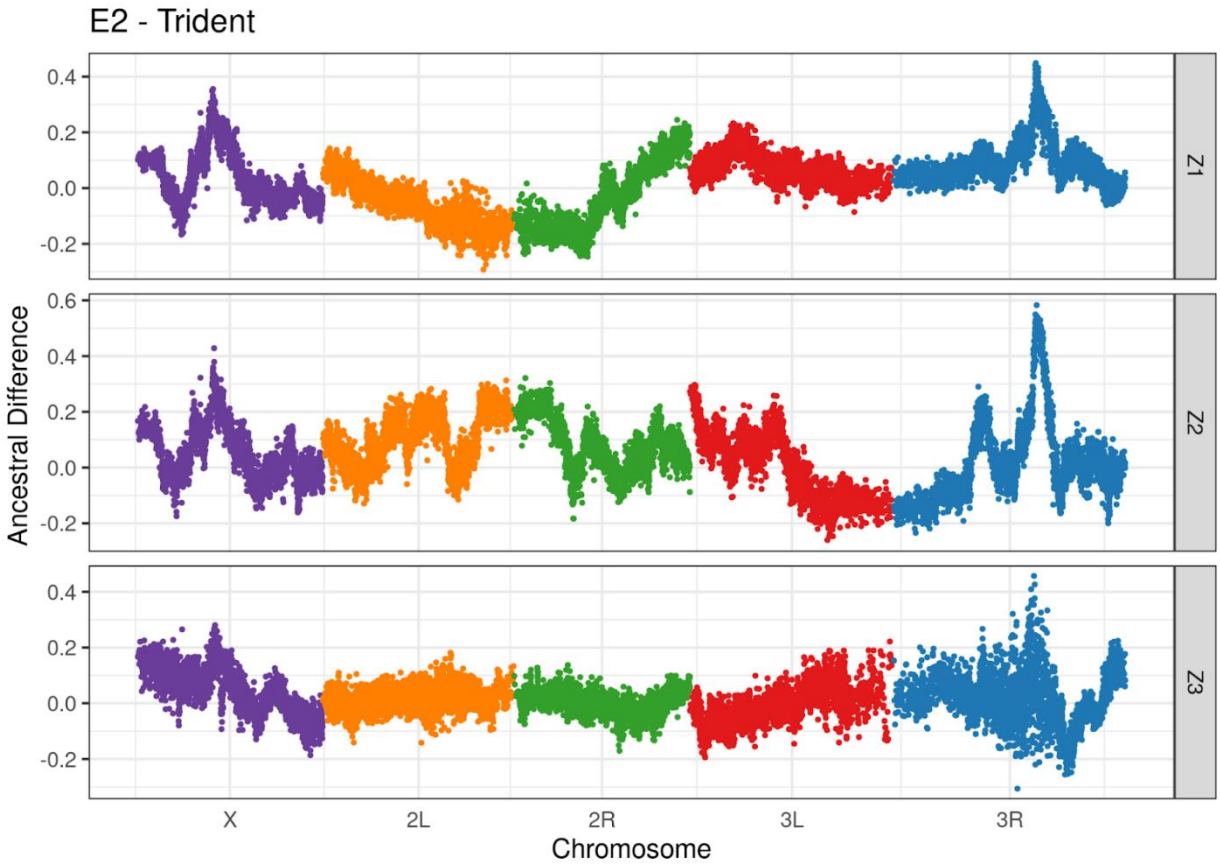
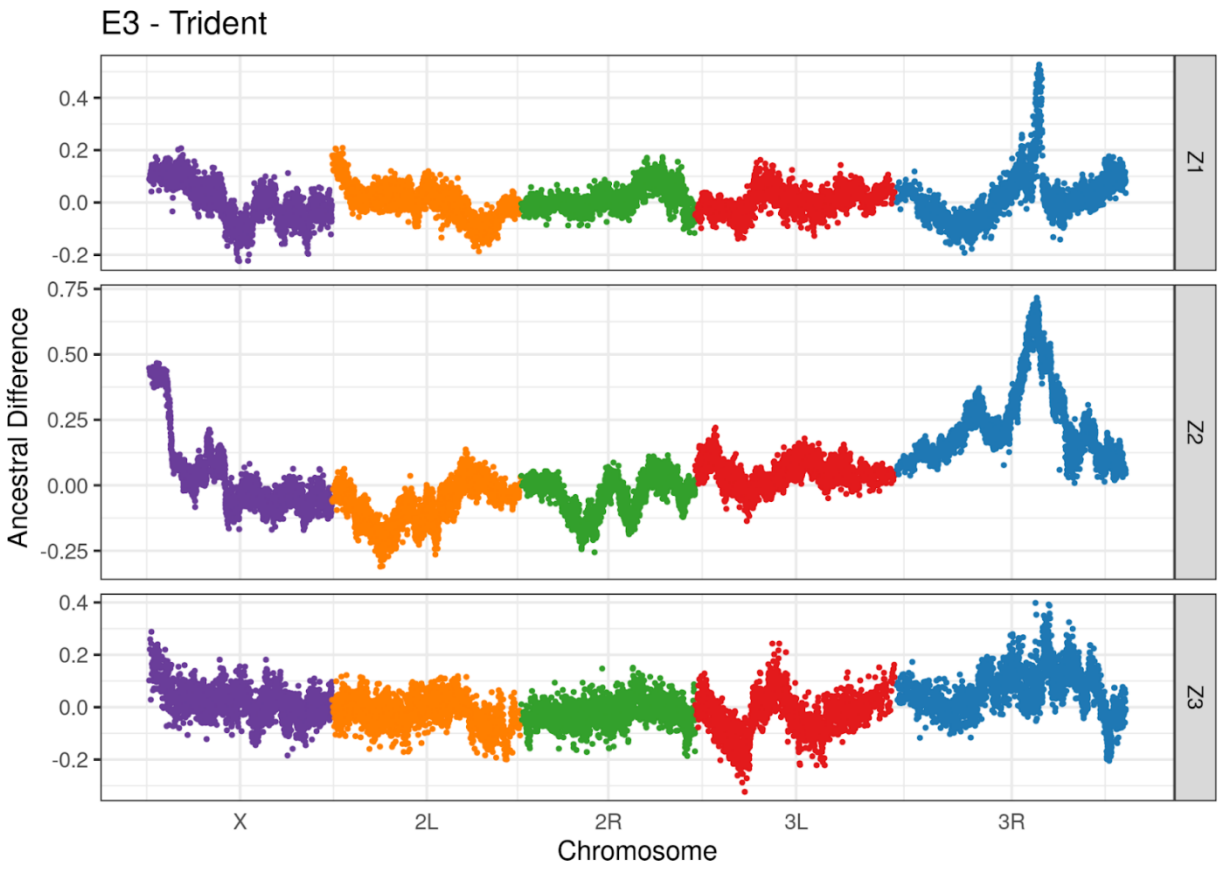
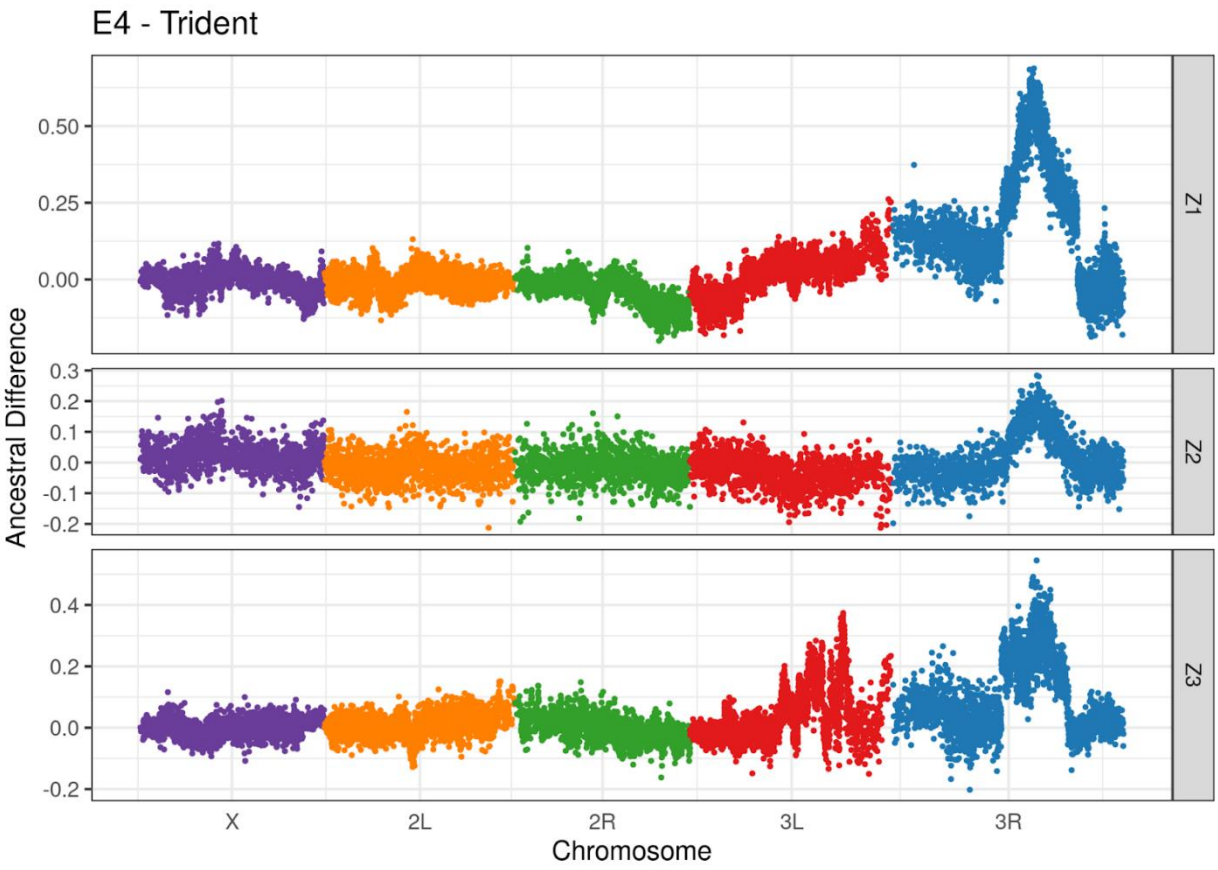


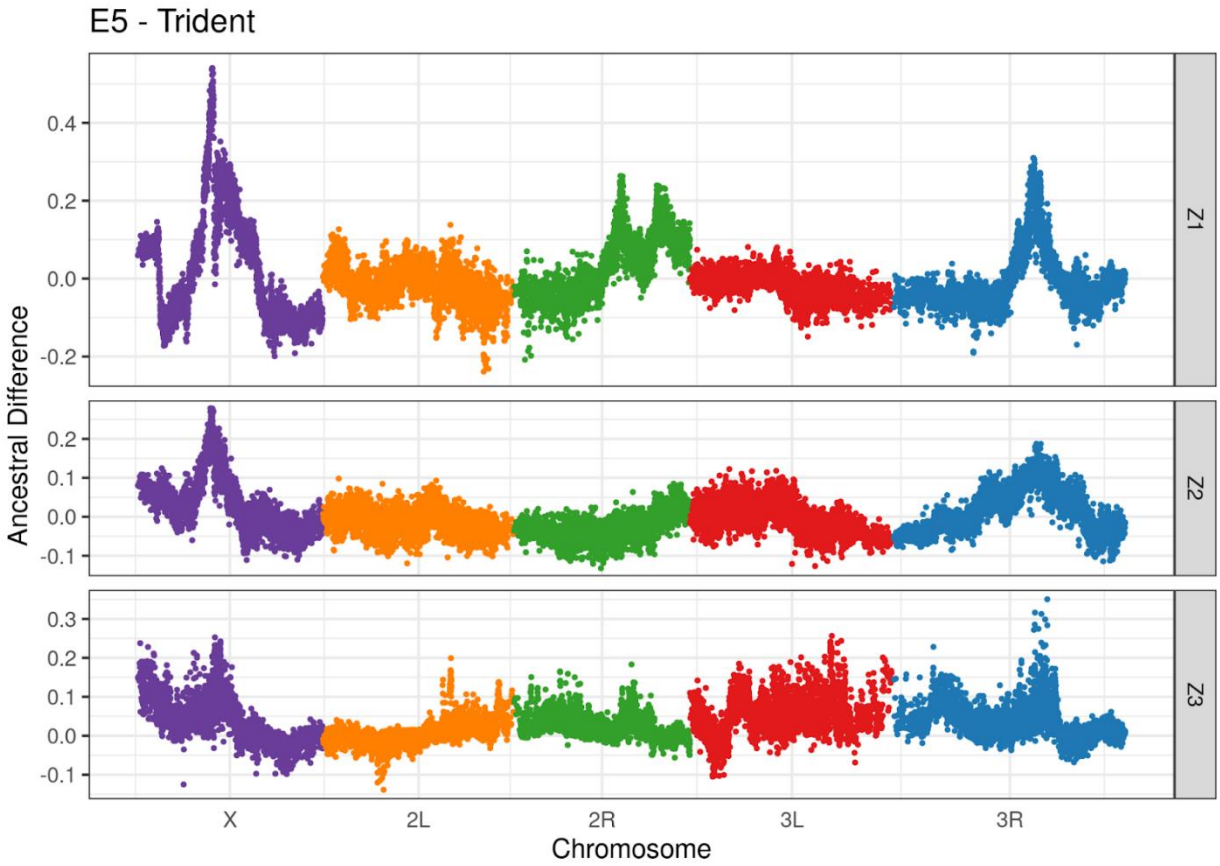
Figure S2. Genomewide ancestry difference distribution for Stripe, different colors representing different chromosome arms. Each page shows three crosses, one Ethiopian by three Zambian parental strains. Ethiopian strain is shown on the top of the plot, and Zambian strains are shown on the right y-axis.

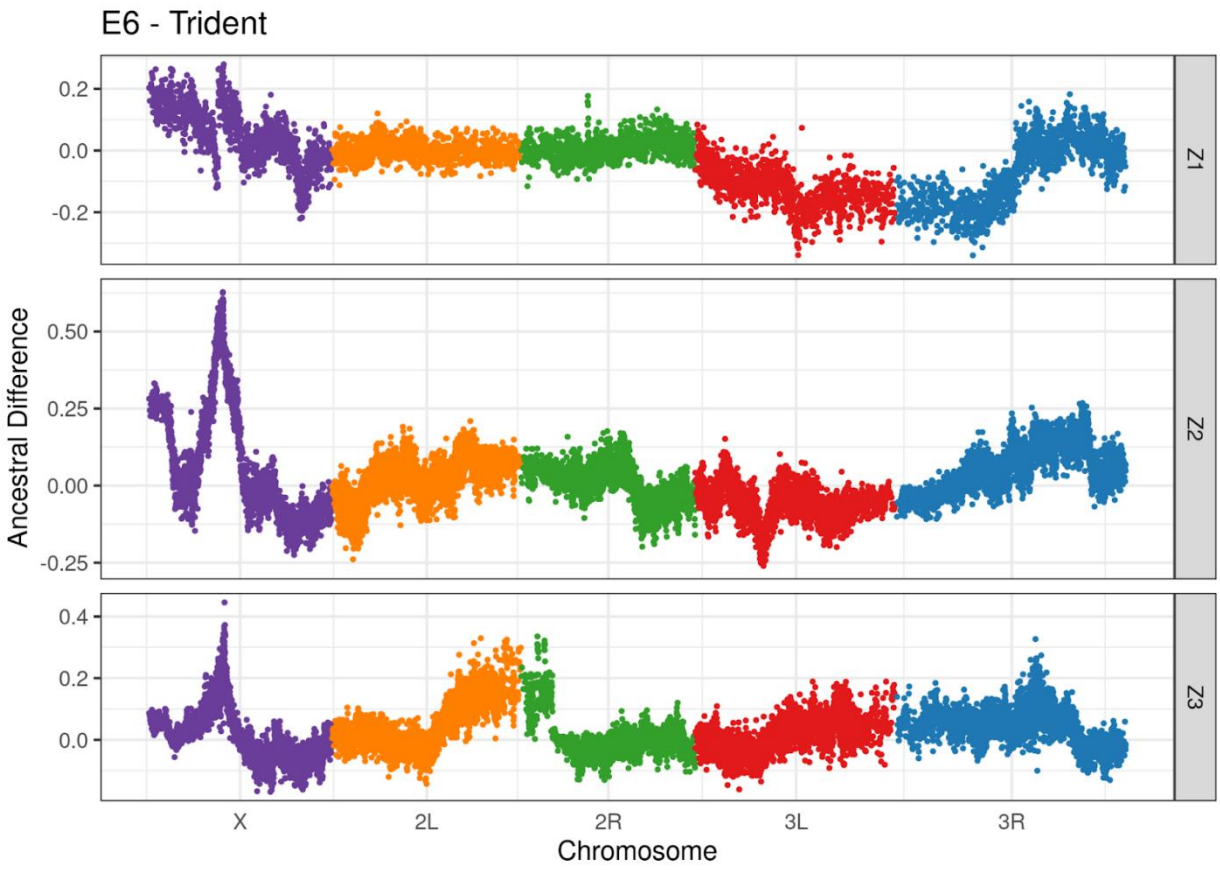

E1 - Trident
Ancestral Difference
0.75
0.50
0.25
0.00
-0.25
Z1
0.50
0.25
0.00
-0.25
Z2
0.4
0.2
0.0
Z3
X
2L
2R
3L
3R
Chromosome


E2 - Trident
Ancestral Difference
0.4
0.2
0.0
-0.2
0.6
0.4
0.2
0.0
-0.2
0.4
0.2
0.0
-0.2
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E3 - Trident
Ancestral Difference
0.4
0.2
0.0
-0.2
0.75
0.50
0.25
0.00
-0.25
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E4 - Trident
Ancestral Difference
0.50
0.25
0.00
0.3
0.2
0.1
0.0
-0.1
-0.2
0.4
0.2
0.0
-0.2
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome


E5 - Trident
Ancestral Difference
0.4
0.2
0.0
-0.2
Z1
0.2
0.1
0.0
-0.1
Z2
0.3
0.2
0.1
0.0
-0.1
Z3
X
2L
2R
3L
3R
Chromosome


E6 - Trident
Ancestral Difference
0.2
0.0
-0.2
0.50
0.25
0.00
-0.25
0.4
0.2
0.0
Z1
Z2
Z3
X
2L
2R
3L
3R
Chromosome

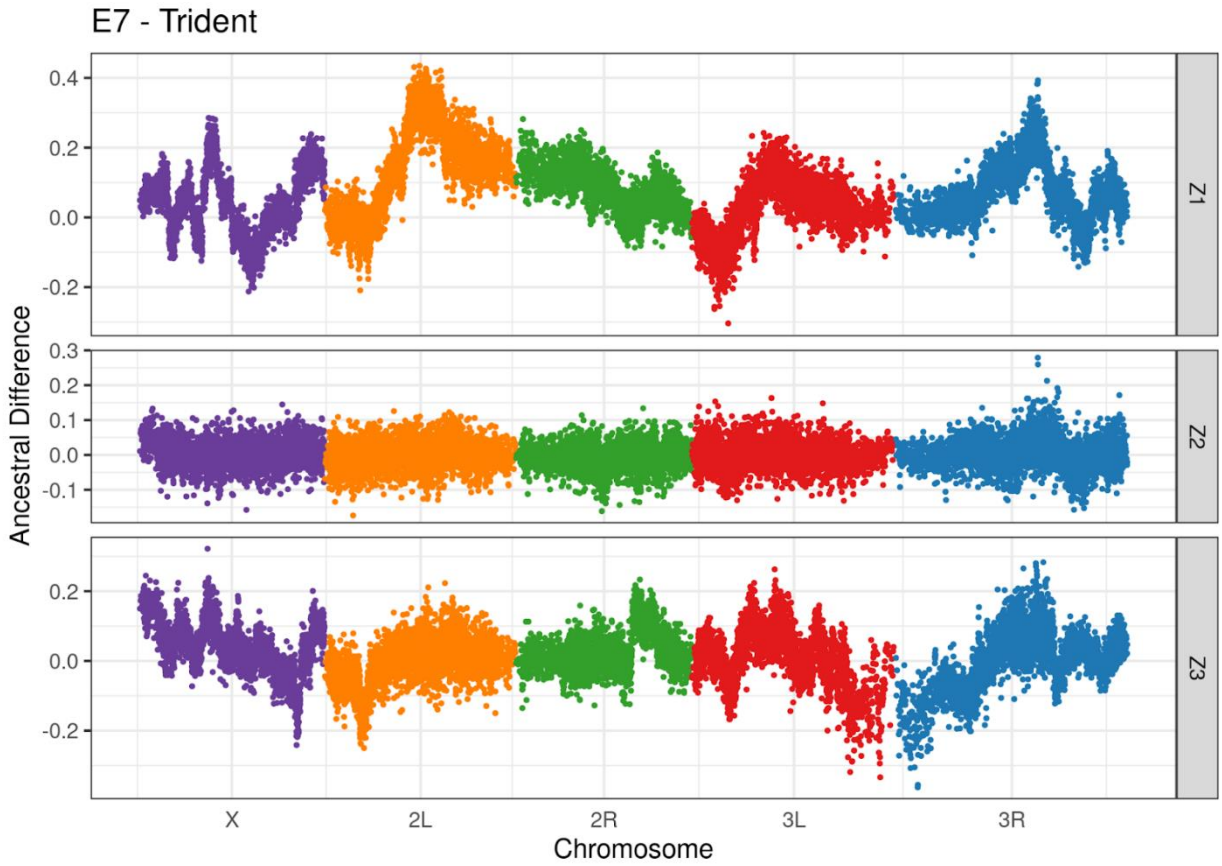


Figure S3. Genomewide ancestry difference distribution for Stripe, different colors representing different chromosome arms. Each page shows three crosses, one Ethiopian by three Zambian parental strains. Ethiopian strain is shown on the top of the plot, and Zambian strains are shown on the right y-axis.

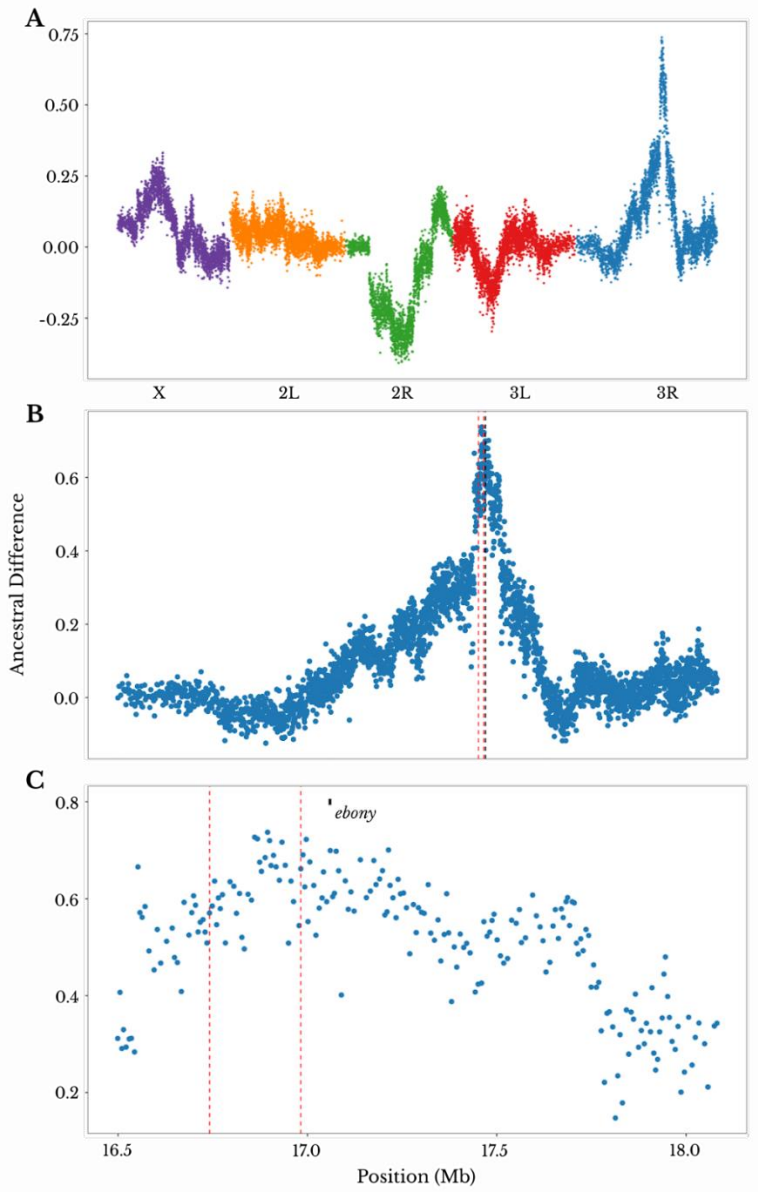


Figure S4. Ancestry distribution for trait mapping Z1E1 Trident shows strong QTL next to *ebony*. (A) Genomewide ancestry difference distribution, different colors representing different chromosome arms. (B) Chromosome arm 3R. (C) QTL with the strongest effect size (0.414) in the experiment, located adjacent to *ebony*, QTL confidence interval shown in

red dashed lines, *ebony* shown as a black bar above the ancestry difference dots.

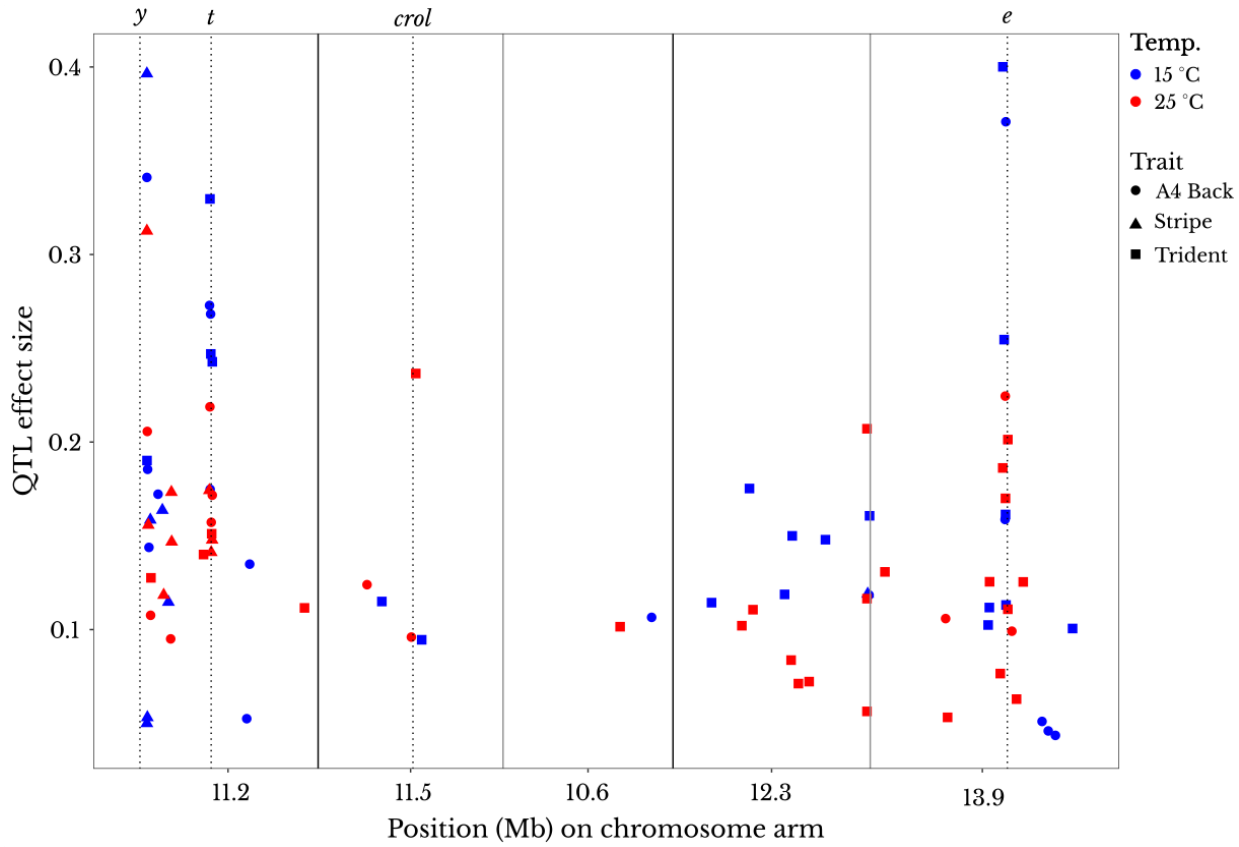


Figure S5. Distribution of effect sizes between cold (blue) and warm (red) QTLs, with shape indicating the different traits.

Chapter 5: Contemporary and ancient genomic signatures of selection in response to salinity transitions in the copepod *Eurytemora affinis* complex (*E. carolleeae*)

Co-author contribution acknowledgment

We would like to thank Martin Bontrager for data collection and fieldwork, as well as assistance in project design, bioinformatics, and data analysis of this project. We also would like to thank Carol E. Lee for her contribution to the project design and writing of this project.

Abstract

Understanding the genetic basis of adaptation on ecological timescales is a pressing issue, as virtually every population is facing the challenges of rapid human-induced environmental changes. However, rapid and recent adaptation to environmental change might involve evolutionary mechanisms that differ from adaptation during ancient events. The *Eurytemora affinis* species complex is a common copepod and a valuable model to study this question, given that it has successfully colonized freshwater habitats several times and is involved in both ancient (~17kya) and contemporary habitat shifts. Thus, we explored the population genomic signatures of selection associated with both ancient and recent salinity transitions by this copepod in the St. Lawrence drainage of North

America, from ancestral higher salinity salt marsh and brackish estuarine habitats to recent freshwater invasions of the Great Lakes. Overall, ion transport was a major biological function under selection across salinities at both ancient and contemporary timescales. We found shared candidate genes under selection across both the ancient transition from salt marsh to brackish estuarine and the recent transition from brackish estuarine to freshwater, including gene paralogs from the *Na^+/K^+-ATPase* and *Na^+/H^+ antiporter* gene families. Between the ancestral saline marsh and brackish estuarine populations, we also found enrichment for genes related to the regulation of transmembrane ion transport. While ion transport was a biological process enriched in all the comparisons, regulation of ion transport was a biological process only enriched in the ancestral habitats, suggesting that ion transport regulation is either a later step that follows the initial colonization of a novel environment or a pre-adaptation that enabled the invasion of fresh water and did not to undergo another round of selection.

Introduction

Understanding the effects of timescale on the genetic basis of adaptive responses is increasingly pressing, as virtually all populations are being challenged by rapid, human-induced changes in the environment. In

contrast to rapid adaptation, the longer a population persists in the same environment, the better adapted it can become, with seemingly no end to how much its fitness can increase (Wiser *et al*. 2013, Lenski *et al*. 2015), but the genetic basis underlying different time points of this process is not necessarily the same. Short- and long-term selective events are likely to differ in the source of adaptive genetic variation and the kinds of mutations being selected. Rapid adaptation has been proposed to rely on standing variation, as new adaptive mutations can take longer to occur and to increase in frequency in a population (Liu *et al*. 1996, Barrett & Schluter 2008, Messer & Petrov 2013). On longer timescales, or at interspecific taxonomic levels and above, selection has been argued to favor less pleiotropic, cis-regulatory mutations (Stern & Orgogozo 2008). Thus, understanding the past evolutionary history of selective pressures acting on a population is critically important for understanding how populations can respond to the challenges that they currently face.

Invasive species are great natural experiments to study rapid adaptation, and further investigating their ancestral ranges can offer useful comparisons to long-term adaptive events. Freshwater invasions, in particular, are likely to impose strong, novel selective pressures on the invading populations, given that a disproportionate number of invaders into freshwater bodies have originated from more saline (mostly brackish) habitats at rates much higher than expectations based on transport

opportunity and propagule pressure (Lee & Bell 1999, Casties *et al.* 2016). Such saline immigrants include some of the most successful invaders in freshwater habitats, including zebra mussels, quagga mussels, and the fishhook water flea (Cristescu *et al*. 2001, Gelembiuk *et al*. 2006, May *et al*. 2006). These brackish populations often inhabit native ranges that are both spatially and temporally heterogeneous in salinity, such that adaptation in the native range to salinity change could predispose them to successfully invade and evolve in response to novel salinities (Lee & Gelembiuk 2008).

The multiple, independent freshwater invasions by the copepod *Eurytemora affinis* complex provide an excellent model for studying the spatial and temporal patterns of adaptation to novel habitats. In particular, the Atlantic clade of this species complex (*E. carolleeae*) encompasses populations in the Great Lakes and throughout the Saint Lawrence estuary. The estuary was colonized thousands of years ago following the end of the Last Glacial Maximum ~17 kya, as the glaciers retreated north and the saltwater populations were able to colonize the brackish habitats (Lee 2000). For thousands of years, the estuarine populations have inhabited different salinities, ranging from hypersaline to brackish salt marshes (~5-40 PSU), and brackish waters near the estuarine turbidity maximum (10-25 PSU) to lower salinity brackish waters closer to the Great Lakes (1-5 PSU). The completely freshwater-invading populations in the Great Lakes, in

turn, have adapted to freshwater habitats in a time span of only ~70 years (Lee 1999, Winkler *et al*. 2008).

Several studies have documented rapid evolutionary responses of *E. affinis* complex populations during the major salinity transitions from brackish to completely freshwater conditions. Such evolutionary shifts include evolutionary changes in freshwater performance and tolerance (Lee *et al*. 2003, Lee *et al*. 2007, 2013), parallel evolutionary shifts in enzymatic activity of the ion transporters *V-type H^+ ATPase* (VHA) and *Na^+/K^+-ATPase* (NKA) (Lee *et al*. 2011), genome-wide evolutionary shifts in gene expression (Posavi *et al*. 2020), and population genomic signatures of natural selection (Stern & Lee, 2020, Stern *et al*. 2022). The genes that show evolutionary shifts in expression and/or population genomic signatures of selection include a suite of ion transport-related genes, such as *Na^+/H^+ antiporter* (*NHA*), *carbonic anhydrase* (*CA*), *VHA*, *NKA*, and *Na^+, K^+, $2Cl^-$ cotransporter* (*NKCC*) (Lee 2021, Posavi *et al*. 2020, Stern & Lee 2020, Stern *et al*. 2022). Our population genomic surveys found signatures of parallel adaptation across independent freshwater invasions from genetically distinct clades, from the St. Lawrence drainage and the Gulf of Mexico (Stern & Lee 2020). Many of the same loci (and SNPs) were under selection across the replicate salinity transitions, including ion transporter gene families, such as *NHA* and *NKA*. In addition, many of the same loci were under selection during laboratory selection imposing salinity decline on a

Baltic Sea population (Stern *et al*. 2022). Additionally, the authors found support for balancing selection in the ancestral saline range as a potential reservoir of standing variation of beneficial alleles during freshwater invasions (Stern & Lee 2020). Given the surmounting evidence of rapid adaptation during freshwater invasions, as well as the existence of populations inhabiting different salinities for thousands of years, the *E. affinis* complex populations of the Saint Lawrence drainage basin offer a great opportunity to study the genetic basis of adaptation at different timescales and salinity gradients.

In this study, we explored the targets of natural selection during the more ancient higher salinity salt marsh to brackish estuarine habitat transition (~17 kya) relative to the loci under selection during the contemporary invasions from brackish estuarine to freshwater habitats (~70 years ago). We aimed to identify the candidate genes and biological functions under selection across the salinity transitions occurring over two different timescales. We specifically aimed to (1) identify candidate regions (genomic windows) and SNPs under selection during freshwater invasions by *E. carolleeae* into two different North American Great Lakes (Lakes Ontario and Michigan), (2) identify candidate regions and SNPs under selection between ancestral habitats in the saline marsh (5-40 PSU) *vs.* brackish estuary (1-5 PSU), and (3) compare the loci under selection during the salinity transitions at the contemporary *versus* ancient time scales.

To achieve these goals, we analyzed four populations of *E. carolleeae* in the St. Lawrence drainage, including two saline populations that colonized the St. Lawrence estuary (~17 kya) after the Last Glacial Maximum (Lee 2000) and two populations that colonized the Great Lakes (~70 years ago) following the opening of the St. Lawrence seaway, around 1958 (Engel 1962). We used SNP frequency data from these populations to search for candidate genes under selection in the saline and freshwater populations and then used Gene Ontology (GO) term enrichment analysis to investigate the biological functions under selection.

Salinity is a major variable structuring aquatic biodiversity on global scales and across all domains of life (Hutchinson 1957, Lozupone & Knight 2007). Rapid evolutionary response to salinity change is an especially salient issue today, given the rapid large-scale salinity declines that are occurring in many parts of the world's oceans due to increased precipitation at high latitudes and the melting of glaciers (Jacobs *et al*. 2002, Nurhati *et al*. 2011, Durack 2015). Here, we studied both populations underlying rapid physiological adaptation on ecological time scales and populations that experienced salinity shifts thousands of years ago. Our results recapitulate the important role of the adaptive evolution of ion transporters in the colonization of freshwater habitats in the *E. affinis* complex and highlight their role in ancient local adaptation to different higher salinity and brackish habitats. The history of natural selection acting

on a population shapes its capacity to evolve in response to future environmental challenges. Thus, the insights gained from this study enhance our ability to understand the capacity of populations to rapidly evolve in response to drastic environmental change and predict future global biodiversity distributions.

Methods

Sampling of copepod populations along a salinity gradient

We sampled ancestral saline and invading freshwater populations of *E. affinis* complex from the genetically distinct Atlantic clade, *E. carolleeae* (Lee 1999, 2000, Alekseev & Souissi 2011, Du *et al. in Review*). Our sampling included two ancestral saline populations in the St. Lawrence estuarine zone and two derived invasive freshwater populations in the Great Lakes of North America (Figure 1). We sampled one ancestral saline population (~5-40 PSU ≈ parts per thousand salinity) from the tidal marsh pools near Baie de L'Isle Verte, QC Canada (lat: 48.00N, long: -69.42W) and another ancestral saline population from the brackish estuary (1-5 PSU) near Montmagny, QC, Canada (46.99N, -70.55W). The two freshwater populations (0 - 0.1 PSU, 350 µS/cm) included one population from Braddock Bay, Lake Ontario (43.31N, -77.71W) and another population from Milwaukee, WI, USA, Lake Michigan (43.05N, -87.88W). From each

sampling location, we isolated 100 adult individuals at an approximately 1:1 ratio of males to females and preserved the animals at -80°C.

Sequencing, alignment, and SNP calling

From each of the four sampled populations (Figure 1), 100 copepods were selected and pooled for DNA extraction and whole-genome sequencing (Pool-seq) (Futschik & Schlötterer 2010). We extracted DNA from each pool of 100 animals using the UltraClean DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) per manufacturer recommendations, with an added 10-minute incubation at 65°C prior to mechanical lysis to increase extraction efficiency. Sequencing libraries were created using the Illumina Nextera DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA). Sequencing was performed on the Illumina HiSeq 2000 platform, using one lane per sample for four samples at the Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine.

After filtering raw reads to exclude low quality sequences and adapter sequences, and trimming low-quality read ends, we aligned the trimmed and filtered reads to the reference genome (using methods described in the next paragraph). The reference genome of the Atlantic clade *E. carolleeae* of the *E. affinis* complex against which we aligned our reads contains 495 Mbp on 6,899 scaffolds (Scaffold 50 = 1,523,809 bp) and was generated from an inbred line derived from a population from the Baie de L'Isle Verte

saltmarsh, Quebec, Canada (Eyun *et al*. 2017). We aligned the reads from each population to the reference genome, obtaining an average depth of coverage of ~22x – 32x per population sample. The population sample from L'Isle Verte, Quebec, Canada (Figure 1) suffered from a lack of sequencing coverage, being 10x lower coverage than the other samples, due to lower initial depth of sequencing coverage.

We used the software package Trimmomatic (Bolger *et al*. 2014) to filter and trim low quality reads, adapter sequences, and unpaired reads from the raw sequencing read data. Repetitive regions of the reference genome were masked with RepeatMasker 4.0.6 (Smit *et al*. 2013) to prevent alignment to those regions, and the filtered, trimmed reads were aligned to the masked reference genome. An initial round of sequence read alignment was first performed with BWA-MEM (Li 2013), followed by a second round of alignment of unaligned reads using Stampy (Lunter & Goodson 2011) to improve the mapping of divergent reads. After alignment, we removed duplicate reads and re-aligned around indels using Picard and GATK IndelRealigner (McKenna *et al*. 2010).

In order to identify SNPs (single nucleotide polymorphisms) between the populations, we processed the read alignments using the SAMtools mpileup utility (Li *et al*. 2009, Li 2011) and the PoPoolation2 mpileup2sync software (Kofler *et al*. 2011) to count variant nucleotides at each position in the reference genome. We removed all positions with

minimum coverage depth lower than 20 reads or within the 2% highest genome-wide coverage depth for any given population. We removed indels and all sites within three positions up- and downstream of the indels to improve the quality of SNP calling. To calculate nucleotide diversity (π), we used all variants from every nucleotide position. To calculate F_{ST}, we calculated minor and major allele frequencies for all positions. For sites with more than two alleles, the minor frequency was the frequency of the second most common SNP allele and the remaining read counts were discarded. The minor allele frequency had to be at least 0.05 across all populations to be considered a SNP.

Detecting genomic signatures of selection

The genomic regions under natural selection have distinct signatures that can be used to distinguish them from regions that are not under selection, including changes in the genetic diversity (such as nucleotide diversity) and the degree of differentiation between populations (such as F_{ST}). To search for genomic signatures of selection between ancestral saline and invading freshwater populations, as well as between the ancestral saline populations, we used all the SNPs that met our quality filtering standards (described above) to estimate pairwise F_{ST} between populations. F_{ST} was calculated following the method of Reynolds *et al*. (1983). We also compared candidate regions under selection between ancestral salt marsh

and brackish estuarine populations to those between the brackish ancestral population and freshwater-derived populations. The goal was to determine whether selection acting between more saline marsh (13-30+ PSU) and brackish estuarine (~1-5 PSU) habitats involved the same genes as those involved in the transition from brackish estuary to freshwater environments.

To detect signatures of selection between the ancestral saline marsh and brackish estuarine populations in the St. Lawrence drainage (Baie de L'Isle Verte *versus* Montmagny) (Figure 1, dark red circles), we calculated the window-wide F_{ST} (F_{ST_Window}) and maximum F_{ST} value for a SNP in a window (F_{ST_MaxSNP}) values between them. The two different F_{ST} approaches were used because they have complementary power in detecting different selective events (da Silva Ribeiro *et al*. 2022), the window-wide approach has higher power to detect selective events in which the adaptive variants are rare when selection starts and selective only raises their frequency moderately, and the maximum SNP approach has a better power to detect instances in which the adaptive variants start at higher frequencies and selection brings them to fixation or near fixation.

Candidate genes under selection between the two ancestral saline populations detected with F_{ST} could have been under selection in either the brackish estuarine or saline marsh habitats, since F_{ST} does not have a directionality and will detect changes that have occurred in either

population. This property contrasts with PBE (Population Branch Excess), an F_{ST}-based method that detects evolutionary events that have happened in only one focal population (Yassin *et al*. 2016). Therefore, pairwise F_{ST} was used to calculate PBE, which compares branch length in one focal population against two other populations and measures the excess of differentiation in the single focal population relative to two other populations (Yassin *et al*. 2016). The PBE statistics is an extension of the Population Branch Statistic (PBS) (Yi *et al*. 2010).

For the PBE analysis, given that we had four populations and were interested in detecting signatures of selection in two freshwater habitats, we used each freshwater population as the focal population in separate PBE analyses and designated the two ancestral saline populations as the background populations in both analyses (Figure 1). For each window, we calculated F_{ST} and PBE for the whole window, as well as for the SNP with the highest F_{ST} within each window. That is, we calculated two different measures of PBE, specifically, (1) PBE_{Window} where we measured F_{ST} for the whole window (F_{ST_Window}), and (2) PBE_{MaxSNP} where we measured F_{ST} for each SNP, and then identified the SNP with the highest PBE value in each window (F_{ST_MaxSNP}).

The *E. affinis* complex genome was divided into 85,332 windows, with an average size of 5,944.36 base pairs per window. For each statistic, we determined the windows in the 99th percentile as outlier windows and,

therefore, candidates to be under selection. To account for events in which nearby windows reflect a single selective sweep with a signature larger than a single window, we combined the windows in the 99th percentile into candidate regions under selection if they were next to each other or separated by no more than five windows. This step accounted for the possibility that the target of selection was present between the two windows, but was not detected due to insufficient coverage. Then, we added 2 kb up- and down-stream of all candidate windows (or merged windows) in our assessments of selection. This step was taken to include genes and regulatory sequences adjacent to the candidate windows, but not necessarily within them.

Gene annotation and GO term enrichment analysis

Gene annotation and GO terms of gene models and transcripts were obtained from Stern and Lee (2020). In brief, automated and whenever possible, manual annotations were used. Manual annotations were previously performed in the Lee Lab (Eyun *et al*. 2017). Gene models were developed mostly based on insect gene identities by the i5K Arthropod Genomes Project dedicated to arthropod genome sequencing (i5K Consortium 2013, Poelchau *et al.* 2014). The transcripts were based on transcriptome sequencing of male and female adults, performed at the Institute for Genome Sciences at the University of Maryland School of

Medicine (Posavi *et al*. 2020). We expanded the list of GO terms associated with each gene to include all the parent terms of each GO term for cellular components, biological processes, and molecular functions.

We performed permutation tests to determine whether our sets of candidate genes, obtained with our outlier analyses, were enriched for a given GO term using a method first described by Pool *et al*. (2012). We counted the number of times a GO term showed up in our list of candidate regions, counting only one occurrence of the GO term per candidate region, and compared it to the number of times it occurred in a random draw of the *E. affinis* complex (Atlantic clade) genome. More specifically, for each set of candidate genomic regions under selection we (1) randomly sampled genomic regions across the *E. affinis* complex genome mimicking the exact number and size of genomic regions in the set of candidate regions empirically detected with our outlier analyses, (2) annotated the genes within the randomized genomic regions, and (3) counted the GO terms obtained for these genes. We repeated this process 10,000 times per permutation test to create a null distribution of the number of counts of each GO term we should expect in a random draw of the *E. affinis* complex genome, given a specific number of genomic regions and their sizes. We considered a GO term significantly enriched if it had a *P*-value lower than 0.01 and had at least 3 counts in our empirical data set.

Results

Candidate genes under selection between the ancestral saltmarsh (L'Isle Verte) and brackish estuarine (Montmagny) populations and between the saline and freshwater invasive (Great Lakes) populations spanned a wide range of functions, mainly related to ion transport, but also amino acid transmembrane transport. One genomic region of particular interest (now known to be on Chromosome 3 in the Atlantic clade genome, Lee 2023, Du *et al. In Review*) contained seven tandem paralogs of the ion transporter gene family *Na^+ /H^+ antiporter* (NHA, SLC9B). Several of the *NHA* paralogs, as well as other genes, have also been shown to exhibit evolutionary shifts in gene expression between saline and freshwater populations in a previous study (Posavi *et al*. 2020) and signatures of selection in wild populations from three genetically distinct clades (sibling species) and in laboratory selection lines (Stern & Lee 2020, Lee 2021, Stern *et al*. 2022).

Biological functions enriched exclusively in the ancestral saline adaptation included the regulation of ion transport. Biological functions enriched in the freshwater adaptation included potassium ion homeostasis. Overall, we found evidence that ion transport has played an important role in the older local adaptation to two different saline habitats (more saline saltmarsh and brackish estuary) and also in the more recent adaptation from brackish to freshwater habitats. The role played by the regulation of

ion transport in the ancestral populations might indicate that given time, further ion transport-related adaptation in freshwater could follow the initial phase of adapting to novel conditions.

DNA sequencing and polymorphism detection

We obtained high-quality genomic data for 129,959,894 base pairs along the *E. carolleeae* draft genome (Eyun *et al*. 2017) (~500 megabases long). We detected 5,406,976 SNPs among the four populations. Median genome-wide nucleotide diversity (π) ranged from 0.021 in the brackish estuarine population at Montmagny, Quebec to 0.015 in the saltmarsh population from Baie de L'Isle Verte, Quebec (Figure 1). Freshwater populations from Lake Ontario (π = 0.021) and Lake Michigan (π = 0.020) had similar levels of π as the ancestral brackish population from Montmagny, indicating no evidence of a population bottleneck following freshwater invasions.

Signatures of selection at ion transmembrane transport genes across the salinity gradient

There was a higher overlap in the genomic regions with signatures of selection in the comparison between the two saline populations (using F_{ST}) than between the two saline populations and each freshwater population (using PBE) (Figure 2). Due to the complementary nature of the statistical

approaches in detecting distinct kinds of selective sweeps (da Silva Ribeiro *et al*. 2022), our lists of candidate genes include genes from both window-wide and maximum SNP statistics (Table S1-S6). We identified 548 candidate genes in genomic regions with signatures of selection between the ancestral saline habitats (salt marsh and brackish estuary, Table S1, S2) and 1,327 candidate genes in genomic regions with signatures of selection in either freshwater habitat (Table S3-S6). Of those, 256 genes showed signatures of selection in both saline and freshwater habitats, including many ion transport-related genes, such as *sodium- and chloride-dependent GABA transporter 1* (Slc6a1), *CA-14*, three *NHA* paralogs (*NHA-4*, *NHA-5,* and *NHA-7*) and two NKA-*α* paralogs (*NKA-α-2*, and *NKA-α-55*) (Table 1, S1-S5). In addition to ion transport function, which was related to the largest enriched GO term (active ion transmembrane transporter activity, GO:0022853) in all but the Lake Michigan PBE_{Window} analyses (Table S7-S11), shared enriched GO terms include amino acid transmembrane transport, which might be related to the transport of osmolytes involved in maintaining cell volume and constant osmotic pressure in and out of the cell, highlighting the selective pressure on transmembrane transport (Table 1).

GO terms seemingly unrelated to ion transport were also enriched in Lake Ontario, such as “bone remodeling and resorption” (GO:0045124). Since *E. affinis* does not have bones, the genes in this category must be

performing other activities, potentially related to calcium transport. Candidate genes under selection in this category included *protein kinase C 1* (PKC1) and *carbonic anhydrase 14* (CA-14). CA-14 has signatures of selection between the ancestral saline populations and in the invasive freshwater populations relative to the ancestral populations (Tables S1-S6) and has previously been implicated in freshwater adaptation across salinity gradients and is a key part of the proposed model for ion uptake in freshwater habitats for the *E. affinis* complex (Posavi *et al*. 2020, Lee 2021, Stern & Lee 2020).

We focused on the genomic region containing seven tandem repeats of the NHA gene family (Scaffold 68, now Chromosome #3 in the new reference genome) and found F_{ST_MaxSNP} outliers in the ancestral saline population comparison and the saline *vs*. freshwater population comparison for both lakes (Table 2). In a region of thirty-nine windows (~146 kb), we found two windows with F_{ST_MaxSNP} outliers in the saline comparison, fourteen PBE_{MaxSNP} outliers in Lake Ontario, and sixteen in Lake Michigan. Two windows were shared between the lakes, but none of the outlier windows between the ancestral saline populations showed signatures of selection in the freshwater populations. Interestingly, the two ancestral saline SNPs were located in exons, one within NHA-4 and the other within NHA-7, and the freshwater SNPs were found in exons, introns, and intergenic regions spanning this region (Table 2). Although part of the

differences could be due to lack of statistical power in detecting signatures of selections in each case, these results suggest that the same gene family, but different paralogs, were used to adapt to different salinities.

Differences between ancestral saline adaption and invasive freshwater adaptation

Focusing on the genes associated with enriched GO terms related to ion transport and ion transport regulation, we can see that 15 genes have signatures of selection only between the two ancestral saline populations, 33 genes have signatures of selection only in the freshwater habitats relative to their ancestral saline, and 22 genes have signatures of selection in both scenarios (Table 3).

Specifically in the adaptation between the two ancestral saline populations, we found an enrichment of genes related to the regulation of ion transport (e.g. GO:2000649, "regulation of sodium ion transmembrane transporter activity"), such as *NHE-X-c* (*sodium/hydrogen exchanger*), *Glycerol-3-phosphate dehydrogenase*, *5-hydroxytryptamine receptor 2B,* and *StAR-related lipid transfer protein 3*. We also found enrichment of genes related to biosynthesis of phosphatidic acid, which is a precursor for the synthesis of other phospholipids, is part of the cell membrane lipid bilayer, and regulates lipid-gated ion channels. Genes in these categories include the *patatin-like phospholipase domain containing 2* (PNPLA2), *glycerol-3-phosphate*

dehydrogenase (Gpdh1), and *retinal degeneration A* (rdgA). Finally, we found enrichment of genes related to polyol (carbohydrate) transport, including the gene *aquaporin 3* (AQP3), which is involved in the transport of nonionic small solutes (Table S7-S8).

In the freshwater lakes, particularly in Lake Michigan, we found enrichment for genes related to ion homeostasis (monovalent inorganic cation homeostasis, GO:0055067, Table S11), a function that relies on ion transport (Dubyak 2004) and involves many of the same genes, but was not enriched in the saline comparison. Unrelated to ion transport, Lake Michigan was enriched for genes involved in rRNA processing, and in Lake Ontario, we found enrichment of genes related to biological functions such as DNA topoisomerase, involved in the winding and unwinding of DNA, and genes related to metabolic processes (Tables S9-S11).

Comparison between the signatures of selection in each freshwater lake

We used the Population Branch Excess (PBE) approach to obtain candidate genomic windows under selection between the two ancestral saline populations and each freshwater lake (Figure 1). We determined PBE for the whole window (PBE_{Window}) and also for the SNP with the highest PBE within each window (PBE_{MaxSNP}). For the PBE_{Window} approach, we obtained 630 candidate regions in Lake Michigan and 688 in Lake Ontario (Table S3, S5). For PBE_{MaxSNP}, we obtained 650 candidate regions in Lake Michigan and 647

in Lake Ontario (Table S4, S6) (Figure 2A and 2B). Within each lake, we found that the overlap between PBE_{Window} and PBE_{MaxSNP} outlier regions ranged from 9.4% to 13.5% (Figure 3). Among the genomic regions detected in both lakes, only 6.3% were detected with both statistics (Figure 3).

We found 18 candidate genes detected with both PBE statistics and in both lakes (Table 4). These candidate genes were predominantly related to ion transport, such as four paralogs of the *Na^+/H^+ antiporter* (NHA, SLC9B) found in tandem in the genome on scaffold 68 (Chromosome 3 in the Atlantic clade) (see below) and two paralogs of *Na^+/K^+-ATPase* (NKA) subunit α (Table 4). Candidate genes under selection detected in only one lake included a larger array of biological functions, including metabolism and homeostasis (complete list of candidate genes detected in each lake with each statistic in Supplementary Tables S3-S6).

Again, we focused on the genomic region containing seven tandem repeats of the NHA gene family, this time investigating every SNP on the top 1% of this genomic region in each lake. In Lake Ontario, we found 22 outlier SNPs, two of which were located in an intronic region of the *neuroendocrine convertase 2* gene (PCSK2), adjacent to the seven tandem paralogs of NHA, another three in intronic regions (NHA-1 and NHA-5), and two in exonic regions (NHA-5 and NHA-6) (Figure 4a, Table S12). In Lake Michigan, we found 33 outlier SNPs, seven in intronic regions (one in PCSK2, one in NHA-1, three in NHA-3, and two in NHA-5) and three in

exonic regions (one in NHA-5 and two in NHA-6, Figure 4b, Table S12). Four outlier SNPs were common in both lakes, in intergenic regions between NHA-6 and NHA-7, and between NHA-5 and NHA-6 (Table S12).

Discussion

The invasion of freshwater habitats by the estuarine and saltmarsh copepod *E. affinis* complex over the last ~70 years required these populations to rapidly adapt to a novel environment (Lee 1999). Perhaps not surprisingly, populations undergoing such a major habitat shift show the evolution of physiological tolerance and performance, as well as the evolution of ion transport activity and expression (Lee 2003, Lee *et al.* 2007, 2011, 2012, Posavi *et al.* 2020). Here, we investigated signatures of selection between two different saline habitats that were colonized thousands of years ago and selection in invasive freshwater populations relative to the ancestral saline populations. The two saline habitats also have different seasonal and daily salinity fluctuation rates, which likely impose different selective pressures on each population. Contrary to the rapid adaptation to freshwater environments during invasions, *E. carolleeae* populations in their native range have had thousands of years to adapt to these different salinities (Winkler *et al.* 2008).

We found that ion transport is a dominant biological function among the candidate gene categories underlying local adaptation between

ancestral saline environments and adaptation from the brackish estuary to freshwater lakes. The most enriched GO term categories in nearly all genome scans for signatures of selection were related to ion transport (Table S7-S11). Despite the difference in salinities and the timescale, we found signatures of selection in ion transport-related genes from the same gene families and in some instances the same gene paralogs (e.g. NHA-4, NHA-5, and NHA-7). The congruence between the genes and gene families under selection during more ancient salinity transitions (between saline and brackish habitats) and during more recent freshwater invasions suggests that the same evolutionary and physiological mechanisms are involved during salinity adaptation across different salinity concentrations (i.e. saline to brackish to fresh) and across different time scales. This result is corroborated by laboratory freshwater selection experiments (Stern *et al*. 2022) and during selection in genetically distinct clades (Stern & Lee 2020, Lee 2021).

Regarding the NHA paralogs, we also showed that additional paralogs (paralogs 1, 3, and 6) were detected under selection only in the invasive freshwater populations relative to the ancestral saline populations. Of these genes, NHA-1, in particular, also showed evolutionary changes in gene expression and was down-regulated in freshwater lines compared to saltwater lines when raised in freshwater conditions (Posavi *et al*. 2020).

Given the previous evidence for the adaptive evolution of ion transporter genes during freshwater invasions by *E. affinis* complex populations, the prevalence of candidate genes related to ion transport followed our expectations. Freshwater habitats contain concentrations of ions orders of magnitude lower than saline habitats (below 1 PSU), making the uptake of ions from the environment much more challenging and energetically costly. Previous studies have shown evolutionary changes related to ion transport during freshwater invasions by the saline copepod *E. affinis* complex (Lee 2021). Parallel signatures of selection across freshwater and saline habitats also involved key ion transporter genes, such as NHA, NKA, and CA (Stern & Lee 2020). Whole transcriptome analysis has shown evolutionary shifts in gene expression between saline and freshwater populations in key ion transport-related genes, including NHA, NKA, carbonic anhydrase, and ammonia transporter (Posavi *et al*. 2020). Our results are congruent with previous studies, and by using metrics designed to capture signatures of local adaptation in several focal populations we add strong evidence to a growing body of literature suggesting that these ion transport genes play a major role in underlying freshwater adaptation (Lee 2021).

The genomic region containing the seven tandem paralogs of the NHA gene was detected in our genome scans for signatures of selection using both PBE_{Window} and PBE_{MaxSNP} metrics. NHA has been discovered in

animals relatively recently (Brett *et al*. 2005, Rheault *et al*. 2007) and most species studied so far have only two paralogs of this gene. The fact that *E. affinis* complex has seven tandem paralogs of NHA (and an eighth paralog in another location of the genome) and that the region with the seven paralogs on scaffold 68 (Chromosome 3) was detected with the PBE_{Window} and PBE_{MaxSNP} approaches suggest that this gene family may be playing an important role in the adaptation to fresh water. Given the close proximity of candidate genomic regions containing the NHA paralog with strong signatures of selection on scaffold 68, it is difficult to pinpoint the exact targets of selection using solely population genomics tools. Thus, future functional assays will be crucial for understanding how and to what extent this region contributes to adaptation. Analyzing the outlier SNPs in this region we found that the two changes in the ancestral saline population comparison were in exons (Table 2), while most freshwater outlier SNPs are located in intergenic or intronic regions, suggesting that the causal mutation under selection might be affecting coding regions in the ancestral population and regulatory regions in the freshwater populations. A causative mutation in the regulatory region in freshwater is congruent with the evolution of the differential expression of NHA-7 between freshwater and saline populations (upregulated in freshwater individuals reared at 15 PSU in both experiments and at 0 PSU in one experiment) (Posavi *et al*. 2020), although we only found a signature of selection on NHA-7 in Lake

Ontario—either due to low power to detect in both lakes or because it is unique to Lake Ontario. We found signatures of selection in both lakes for NHA-6 and NHA-*5*, of which only NHA-*5* has evolved differences in gene expression (down-regulated in freshwater individuals reared at 0 PSU in one experiment). Given the high number of outlier SNPs in this region and the fact that two NHA paralogs have shown evolutionary shifts in gene expression, we cannot rule out that more than one SNP might be the target of selection in this region. These results are congruent with broad parallel signatures of selection between different lineages of *E. affinis* complex (Stern & Lee 2020, Lee 2021).

Ancient, local adaptation to different salinities also seems to have uniquely invoked genes related to the regulation of ion transport itself, strengthening the hypothesis that this is a crucial biological function to survive at different salinities (Table 3). This result raises the question of whether ion transport is a key first step in the evolution of freshwater habitats and will be followed by additional changes that will continuously increase the population's fit to its environment. In part, selection in freshwater invasive populations might have used a subset of the genes selected at older timescales, as seen in insulin-like receptor evolution in *Drosophila* (Guira-Rico & Aguadé 2009).

Conclusions

Our study sheds light on the tempo and mode of adaptive change during habitat transitions. Our results are congruent with a scenario in which adaptation at different salinity and timescales acted on ion transport, while selection in a longer timescale also acted on ion transport regulation, suggesting that more opportunities for genetic variants that could fine-tune major biological functions such as ion transport could be selected and consequently increases the fitness of the population. Comparing adaptive events in geological and ecological timescales and understanding their particularities is especially relevant today, as populations that have had thousands of years to adapt to their habitats face the threats of human-induced rapid environmental changes.

Acknowledgments

This work was supported by the National Science Foundation OCE-1658517, NSF DEB-2055356, & French National Research Agency ANR-19-MPGA-0004 to Carol E. Lee. Tiago da Silva Ribeiro was supported by a UW-Madison Graduate Recruitment Fellowship and UW-Madison Department of Integrative Biology Graduate Summer Research Awards. Martin Bontrager was supported by an NSF Predoctoral Fellowship and NIH funded UW-Genetics training grant.

References

Alekseev, V., & Souissi, A. (2011). A new species within the *Eurytemora affinis* complex (Copepoda: Calanoida) from the Atlantic Coast of USA, with observations on eight morphologically different European populations. *Zootaxa*, *2767*, 41–56. https://doi.org/10.11646/zootaxa.2767.1.4.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38-44. https://doi.org/10.1016/j.tree.2007.09.008.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brett, C. L., Donowitz, M., & Rao, R. (2005). Evolutionary origins of eukaryotic sodium/proton exchangers. *American Journal of Physiology-Cell Physiology*, *288*(2), C223–C239. https://doi.org/10.1152/ajpcell.00360.2004.

Casties, I., Seebens, H., & Briski, E. (2016). Importance of geographic origin for invasion success: A case study of the North and Baltic Seas versus the Great Lakes–St. Lawrence River region. *Ecology and Evolution*, *6*(22), 8318–8329. https://doi.org/10.1002/ece3.2528.

Cristescu, M. E. A., Hebert, P. D. N., Witt, J. D. S., MacIsaac, H. J., & Grigorovich, I. A. (2001). An invasion history for *Cercopagis pengoi* based on mitochondrial gene sequences. *Limnology and Oceanography*, *46*(2), 224–229. https://doi.org/10.4319/lo.2001.46.2.0224.

da Silva Ribeiro, T., Galván, J. A., & Pool, J. E. (2022). Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation. *Genome Biology and Evolution*, 14(10), evac143. https://doi.org/10.1093/gbe/evac143.

Du, Z., Gelembiuk, G., Moss, W., Tritt, A., & Lee, C. E. (*In Review*). The genome architecture of a copepod invading novel habitats. *Genomics, Proteomics & Bioinformatics.*

Dubyak, G. R. (2004). Ion homeostasis, channels, and transporters: an update on cellular mechanisms. *Advances in Physiology Education*, 28(4), 143-154. https://doi.org/10.1152/advan.00046.2004.

Durack, P. J. (2015). Ocean salinity and the global water cycle. *Oceanography*, *28*(1), 20–31.

Engel, R. A. (1962). *Eurytemora sffinis*, a calanoid copepod new to Lake Erie. *Ohio Journal of Science*, *62*, 252.

Eyun, S., Soh, H. Y., Posavi, M., Munro, J. B., Hughes, D. S. T., Murali, S. C., Qu, J., Dugan, S., Lee, S. L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K. C., Muzny, D. M., Park, E.-O., Silva, J. C., Gibbs, R. A., Richards, S., & Lee, C. E. (2017). Evolutionary history of chemosensory-related gene families across the Arthropoda. *Molecular Biology and Evolution*, *34*(8), 1838–1862. https://doi.org/10.1093/molbev/msx147.

Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA Samples. *Genetics*, *186*(1), 207–218. https://doi.org/10.1534/genetics.110.114397.

Gelembiuk, G. W., May, G. E., & Lee, C. E. (2006). Phylogeography and systematics of zebra mussels and related species: Phylogeography of zebra mussels. *Molecular Ecology*, *15*(4), 1033–1050. https://doi.org/10.1111/j.1365-294X.2006.02816.x.

Guirao-Rico, S., & Aguadé, M. (2009). Positive selection has driven the evolution of the *Drosophila* insulin-like receptor (*InR*) at different timescales. *Molecular Biology and Evolution*, 26(8), 1723-1732. https://doi.org/10.1093/molbev/msp088.

Hutchinson, G. E. (1957). *A Treatise on Limnology* (1st ed.). John Wiley & Sons, Inc.

i5K Consortium. (2013). The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), 595-600. https://doi.org/10.1093/jhered/est050.

Jacobs, S. S., Giulivi, C. F., & Mele, P. A. (2002). Freshening of the Ross Sea during the late 20th century. *Science*. https://doi.org/10.1126/science.1069574.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589.

Lee, C. E. (1999). Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, *53*(5), 1423–1434. https://doi.org/10.2307/2640889.

Lee, C. E. (2000). Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations." *Evolution*, *54*(6), 2014–2027. https://doi.org/10.1111/j.0014-3820.2000.tb01245.x.

Lee, C. E. (2003). Evolution of physiological tolerance and performance during freshwater invasions. *Integrative and Comparative Biology*, *43*(3), 439–449. https://doi.org/10.1093/icb/43.3.439.

Lee, C. E. (2021). Ion transporter gene families as physiological targets of natural selection during salinity transitions in a copepod. *Physiology*, 36(6), 335-349. https://doi.org/10.1152/physiol.00009.2021.

Lee, C. E. (2023). Genome architecture underlying salinity adaptation in the invasive copepod *Eurytemora affinis* species complex: A review. *Iscience*. doi:10.1016/j.isci.2023.107851.

Lee, C. E., & Bell, M. A. (1999). Causes and consequences of recent freshwater invasions by saltwater animals. *Trends in Ecology & Evolution*, *14*(7), 284–288. https://doi.org/10.1016/S0169-5347(99)01596-7.

Lee, C. E., & Gelembiuk, G. W. (2008). Evolutionary origins of invasive populations. *Evolutionary Applications*, 1(3), 427-448. https://doi.org/10.1111/j.1752-4571.2008.00039.x.

Lee, C. E., Kiergaard, M., Gelembiuk, G. W., Eads, B. D., & Posavi, M. (2011). Pumping ions: Rapid parallel evolution of ionic regulation following habitat invasions. *Evolution*, *65*(8), 2229–2244. https://doi.org/10.1111/j.1558-5646.2011.01308.x.

Lee, C. E., Moss, W. E., Olson, N., Chau, K. F., Chang, Y., & Johnson, K. E. (2013). Feasting in fresh water: Impacts of food concentration on freshwater tolerance and the evolution of food × salinity response during the expansion from saline into fresh water habitats. *Evolutionary Applications*, *6*(4), 673–689. https://doi.org/10.1111/eva.12054.

Lee, C. E., Posavi, M., & Charmantier, G. (2012). Rapid evolution of body fluid regulation following independent invasions into freshwater

habitats. *Journal of Evolutionary Biology*, *25*(4), 625–633. https://doi.org/10.1111/j.1420-9101.2012.02459.x.

Lee, C. E., Remfert, J. L., & Chang, Y.-M. (2007). Response to selection and evolvability of invasive populations. *Genetica*, *129*(2), 179–192. https://doi.org/10.1007/s10709-006-9013-9.

Lenski, R. E., Wiser, M. J., Ribeck, N., Blount, Z. D., Nahum, J. R., Morris, J. J., Zaman, L., Turner, C. B., Wade, B. D., Maddamsetti, R., Burmeister, A. R., Baird, E. J., Bundy, J., Grant, N. A., Card, K. J., Rowles, M., Weatherspoon, K., Papoulis, S. E., Sullivan, R., Clark, C., Mulka, J. S., & Hajela, N. (2015). Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821), 20152292. https://doi.org/10.1098/rspb.2015.2292.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Liu, J., Mercer, J. M., Stam, L. F., Gibson, G. C., Zeng, Z. B., & Laurie, C. C. (1996). Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics*, 142(4), 1129-1145. https://doi.org/10.1093%2Fgenetics%2F142.4.1129.

Lozupone, C. A., & Knight, R. (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences*, *104*(27), 11436–11440. https://doi.org/10.1073/pnas.0611525104.

Lunter, G., & Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, *21*(6), 936–939. https://doi.org/10.1101/gr.111120.110.

May, G. E., Gelembiuk, G. W., Panov, V. E., Orlova, M. I., & Lee, C. E. (2006). Molecular ecology of zebra mussel invasions: zebra mussel invasions. *Molecular Ecology*, *15*(4), 1021–1031. https://doi.org/10.1111/j.1365-294X.2006.02814.x.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 28(11), 659-669. https://doi.org/10.1016/j.tree.2013.08.003.

Nurhati, I. S., Cobb, K. M., & Lorenzo, E. D. (2011). Decadal-scale SST and salinity variations in the central tropical pacific: Signatures of natural and anthropogenic climate change. *Journal of Climate*, *24*(13), 3294–3308. https://doi.org/10.1175/2011JCLI3852.1.

Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., Lin, H., Lin, J. W., & Hackett, K. (2015). The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714-D719. https://doi.org/10.1093/nar/gku983.

Pool, J. E., Corbett-Detig, R. B., Sugino, R. P., Stevens, K. A., Cardeno, C. M., Crepeau, M. W., Duchen, P., Emerson, J. J., Saelao, P., Begun, D. J., & Langley, C. H. (2012). Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLOS Genetics*, *8*(12), e1003080. https://doi.org/10.1371/journal.pgen.1003080.

Posavi, M., Gulisija, D., Munro, J. B., Silva, J. C., & Lee, C. E. (2020). Rapid evolution of genome-wide gene expression and plasticity during saline to freshwater invasions by the copepod *Eurytemora affinis* species complex. *Molecular Ecology*, *29*(24), 4835–4856. https://doi.org/10.1111/mec.15681.

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, *105*(3), 767–779. https://doi.org/10.1093/genetics/105.3.767.

Rheault, M. R., Okech, B. A., Keen, S. B. W., Miller, M. M., Meleshkevitch, E. A., Linser, P. J., Boudko, D. Y., & Harvey, W. R. (2007). Molecular

cloning, phylogeny and localization of *AgNHA1*: The first *Na+/H+ antiporter* (*NHA*) from a metazoan, *Anopheles gambiae*. *Journal of Experimental Biology*, *210*(21), 3848–3861. https://doi.org/10.1242/jeb.007872.

Smit, Arian F, Hubley, Robert, & Green, P. (2013). *RepeatMasker Open-4.0*. http://www.repeatmasker.org.

Stern, D. B., Anderson, N. W., Diaz, J. A., & Lee, C. E. (2022). Genome-wide signatures of synergistic epistasis during parallel adaptation in a Baltic Sea copepod. *Nature Communications*, 13(1), 4024. https://doi.org/10.1038/s41467-022-31622-8.

Stern, D. B., & Lee, C. E. (2020). Evolutionary origins of genomic adaptations in an invasive copepod. *Nature Ecology & Evolution*, *4*(8), 1084–1094. https://doi.org/10.1038/s41559-020-1201-y.

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution*, 62(9), 2155-2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x.

Winkler, G., Dodson, J. J., & Lee, C. E. (2008). Heterogeneity within the native range: Population genetic analyses of sympatric invasive and noninvasive clades of the freshwater invading copepod *Eurytemora affinis*. *Molecular Ecology*, *17*(1), 415–430. https://doi.org/10.1111/j.1365-294X.2007.03480.x.

Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, *342*(6164), 1364-1367. https://doi.org/10.1126/science.1243357.

Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., David, J. R., & Pool, J. E. (2016). Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proceedings of the National Academy of Sciences*, *113*(17), 4771–4776. https://doi.org/10.1073/pnas.1522559113.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T. S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., … Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75–78. https://doi.org/10.1126/science.1190371.

Table 1. Shared gene ontology (GO) terms enriched in both the ancestral saline and the invasive freshwater genome scans for signatures of selection. The category indicates whether the GO term represents a biological process (b), cellular component (c), or molecular function (m). The genes and *P*-values (all P-values < 0.01) associated with each GO term for different analyses are shown in Tables S7-S11.

GO Term	**Category**	**Name**
GO:1902600	b	proton transmembrane transport
GO:1902475	b	L-α-amino acid transmembrane transport
GO:0045989	b	positive regulation of striated muscle contraction
GO:0002026	b	regulation of the force of heart contraction
GO:0098533	c	ATPase dependent transmembrane transport complex
GO:0090533	c	cation-transporting ATPase complex
GO:0022853	m	active ion transmembrane transporter activity
GO:0015491	m	cation:cation antiporter activity
GO:0015298	m	solute:cation antiporter activity
GO:0015297	m	antiporter activity
GO:0015291	m	secondary active transmembrane transporter activity
GO:0015179	m	L-amino acid transmembrane transporter activity

Table 2. Top 1% outlier for ancestral saline F_{ST_MaxSNP} analysis and invasive freshwater PBE_{MaxSNP} analyses within the genomic region of scaffold 68 containing seven tandem repeats of the NHA gene. "Position" refers to the genomic coordinates in the i5K *E. carolleeae* reference genome, located at the i5K workspace (https://i5k.nal.usda.gov/Eurytemora_affinis).

Position	Saltwater *vs.* Brackish	Lake Ontario	Lake Michigan	Genomic region	Gene
534149			X	Intron	PCSK2
534322		X		Intron	PCSK2
536985		X		Intron	PCSK2
547744	X			Exon	NHA-7
561936		X	X	Intergenic	NHA-7 - NHA-6
565628		X		Intergenic	NHA-7 - NHA-6
565712			X	Intergenic	NHA-7 - NHA-6
567091			X	Intergenic	NHA-7 - NHA-6
567938		X		Intergenic	NHA-7 - NHA-6
579255			X	Exon	NHA-6
580509		X		Exon	NHA-6
581289			X	Exon	NHA-6
586351		X		Intergenic	NHA-6 - NHA-5
586437		X	X	Intergenic	NHA-6 - NHA-5
590889			X	Exon	NHA-5
590931		X		Exon	NHA-5

593323			X	Intron	NHA-5
601415	X			Exon	NHA-4
622703		X		Intergenic	NHA-4 - NHA-3
624753			X	Intergenic	NHA-4 - NHA-3
636388			X	Intron	NHA-3
644702		X		Intergenic	NHA-3 - NHA-2
648304			X	Intergenic	NHA-3 - NHA-2
657560			X	Intergenic	NHA-2 - NHA-1
662722		X		Intergenic	NHA-2 - NHA-1
663185			X	Intergenic	NHA-2 - NHA-1
666245		X		Intergenic	NHA-2 - NHA-1
666472			X	Intergenic	NHA-2 - NHA-1
671662		X		Intron	NHA-1
674159			X	Intergenic	NHA-1 -

Table 3. Ion transport-related genes under selection only between the two ancestral saline populations (column 1), only in the freshwater lakes relative to ancestral saline (column 3), and both (column 2). Gene descriptions can be seen in supplementary materials (Table S1-S6). GO terms associated with these genes can be seen in supplementary materials (Tables S7-S11).

Ancestral Saline	Both	Invasive Freshwater
5_HT1B	EAFF012604	EAFF008247
At3g05155	EaffTmpM007749	EAFF008536
Ca_P60A	NBC	EAFF008893
EAFF017857	NHA-5	EAFF009605
EAFF025707	NHA-7	EAFF009606
Gpdh1	NKA-a-2	Gid4
mec_2	NKA-a-5	Indy
MRTO4	pbo_4	KCNJ18
nac_1	SFXN1	KCNJ2
NHE-X-c	Slc13a2	nAChRbeta2
NKCC-frag	Slc13a3	Nckx30C
SLC4A8	Slc2a1	NHA-1
slc5a9	SLC2A13	NHA-3
SLC6A13	Slc6a1	NHA-6
sto_2	Slc6a18	NHE2_5
	Slc6a5	NKA-a-1
	VhaAC39_1	NKA-b-4
	NHA-4	Orct
	Nach-PPK28	SLC12A6
	Stard3	SLC13A5
	Tret1	Slc20a1
	Tret1_2	SLC22A5
		SLC26A11
		Slc36a2
		Slc45a2
		SLC4A10
		SLC4A11
		Slc6a7
		SLC8A1

		Slc9a9
		slo
		surf1
		w

Table 4. List of shared candidate genes in both Lake Michigan and Lake Ontario that were detected with both PBE_{Window} and PBE_{MaxSNP}. *EaffTmpM* indicates a temporary gene code.

Gene ID	**Gene Symbol**	**Description**
XLOC_036449	*setd7*	*Histone-lysine N-methyltransferase SETD7*
XLOC_035875	*PDCD11*	*Protein RRP5 homolog*
XLOC_035467	*KCNJ18*	*Inward rectifier potassium channel 18*
XLOC_034641	*EaffTmpM010522*	*chromaffin granule amine transporter, putative*
XLOC_032688	*NHA-1*	*Na+/H+ Antiporter, paralog 1*
XLOC_032687	*NHA-1-frag*	*Na+/H+ Antiporter, paralog 1 (fragment)*
XLOC_032685	*NHA-3*	*Na+/H+ Antiporter, paralog 3*
XLOC_032683	*NHA-5*	*Na+/H+ Antiporter, paralog 5*
XLOC_032682	*NHA-6*	*Na+/H+ Antiporter, paralog 6*
XLOC_032632	*CHRNA7*	*Neuronal acetylcholine receptor subunit alpha-7*
XLOC_032620	*PCSK2*	*Neuroendocrine convertase 2*
XLOC_028276	*sll0108*	*Putative ammonium transporter sll0108*
XLOC_022586	*MANBA*	*Beta-mannosidase*
XLOC_020867	*NKA-a-1*	*Na+/K+-ATPase, subunit alpha, paralog 1*
XLOC_018666	*EaffTmpM006183*	*hypothetical protein GUITHDRAFT 82324*
XLOC_015897	*Mkx*	*Homeobox protein Mohawk*
XLOC_014637	*NKA-a-2*	*Na+/K+-ATPase, subunit alpha, paralog 2*
XLOC_014506	*EaffTmpM022651*	*EF-hand domain-containing protein D1*

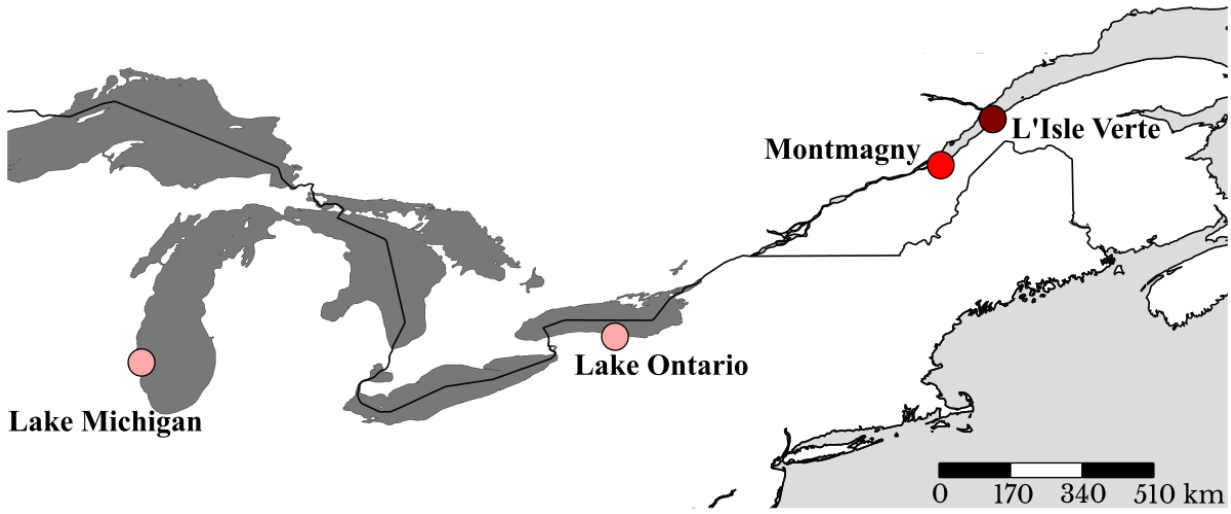


Figure 1. Sampling of populations of *Eurytemora carolleeae*, from the Atlantic clade of the *E.affinis* complex for this study. Light red circles: freshwater populations (salinity range: 0-0.1‰) from Lake Michigan and Lake Ontario. Dark red circles: the brackish population from Montmagny (salinity range: 1-5‰) and the saltwater population from Baie de L'Isle Verte (salinity range: 13-40‰).

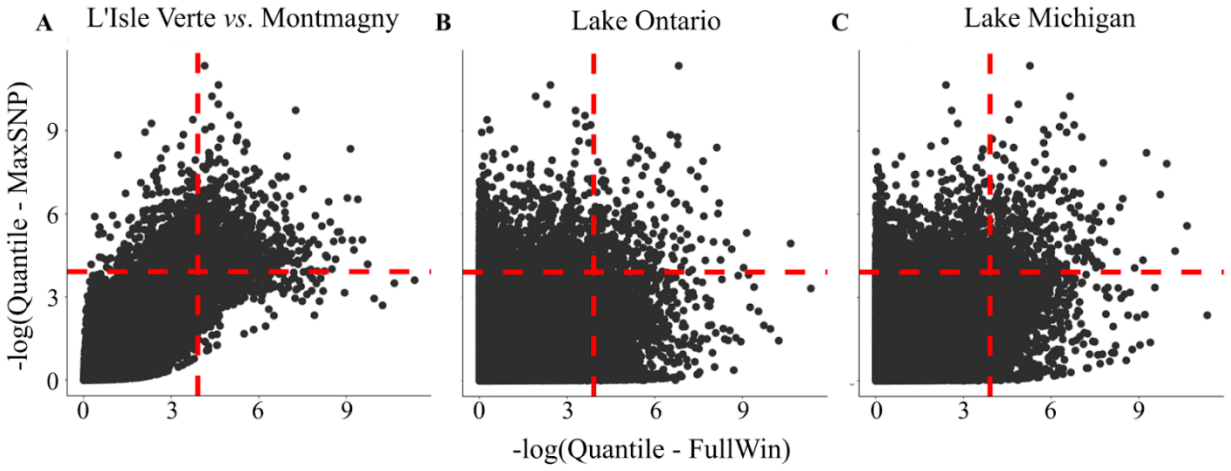


Figure 2. Overlap in outlier regions between the window-wide and highest SNP statistics. (A) L'Isle Verte (saltmarsh) *vs*. Montmagny (brackish) ancestral saline habitats using F_{ST} analysis, (B) Lake Ontario and (C) Lake Michigan *vs*. L'Isle Verte and Montmagny using PBE analyses. The upper right quadrants show the overlap in outlier regions between the two statistics. Dashed lines represent the 99th percentile cutoff for each statistic. Each dot represents the F_{ST} or PBE value in a window across the genome.

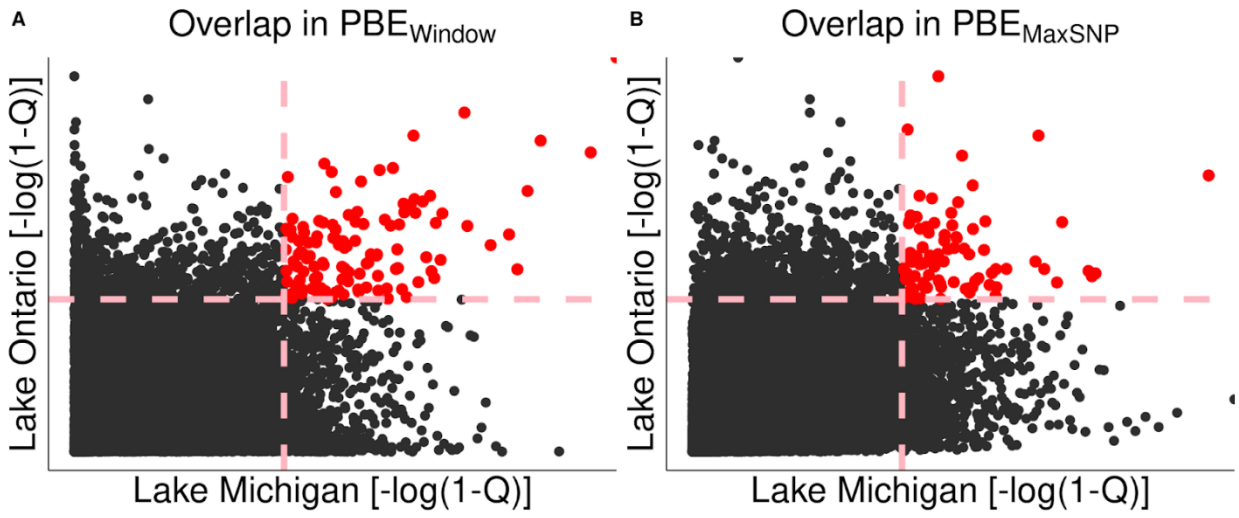


Figure 3. Overlap in outlier regions between the Lake Michigan and Lake Ontario populations using: (A) PBE_{Window} from each lake relative to both saline populations, and (B) PBE_{MaxSNP} from each lake relative to both saline populations. The upper right quadrants (red dots) show the overlap in

outlier regions between the Lake Michigan and Lake Ontario populations. Dashed lines represent the 99th percentile cutoff for each lake. Each dot represents the PBE value in a window across the genome.

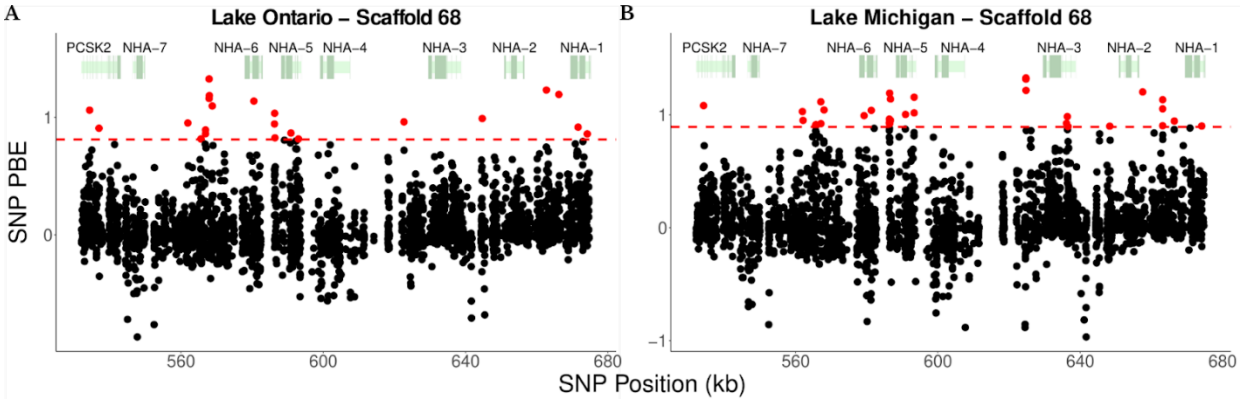


Figure 4. Distribution of PBE values for each individual SNP in a candidate region within Scaffold 68 (chromosome 3) containing seven paralogs of the NHA gene family for (a) Lake Ontario and (b) Lake Michigan. The dashed red line indicates the genome-wide PBE value cutoff for the 1% outliers for PBE_{MaxSNP}. Red dots indicate outlier SNPs above the cutoff. Dark green thick bars show the exonic regions of the NHA paralogs, whereas the light green thin bars indicate the intronic regions of the genes.

Supplementary tables and figures

Table S1. Candidate genes under selection in Montmagny or L'Isle Verte, the ancestral saline habitats, detected with F_{ST_Window}.

Gene ID	Gene Symbol	Description
XLOC_000592	Rh-2	Rh protein, paralog 2
XLOC_000707	At4g35335	CMP-sialic acid transporter 4
XLOC_000808	EaffTmpM029332	luciferase
XLOC_001043	EaffTmpM012603	hypothetical protein
XLOC_001052	EaffTmpM012607	Serine-aspartate repeat-containing protein I
XLOC_001055	CA-14	Carbonic Anhydrase, paralog 14
XLOC_001058	ACAD10	Acyl-CoA dehydrogenase family member 10
XLOC_001063	Tdrd3	Tudor domain-containing protein 3
XLOC_001068	SFXN1	Sideroflexin-1
XLOC_001083	unc45b	Ribonuclease P protein subunit rpr2
XLOC_001095	MYLK	Myosin light chain kinase, smooth muscle
XLOC_001099	acsbg2	Long-chain-fatty-acid-CoA ligase ACSBG2
XLOC_001100	add	Adenosine deaminase
XLOC_001104	rho 5	Inactive rhomboid protein 1
XLOC_001105	EaffTmpM012611	RNA-binding protein 12B
XLOC_001115	VDE1	Violaxanthin de-epoxidase, chloroplastic
XLOC_001120	YAP1	Transcriptional coactivator YAP1
XLOC_001129	Sumo3	Small ubiquitin-related modifier 3
XLOC_001132	NAXD	ATP-dependent (S)-NAD(P)H-hydrate dehydratase
XLOC_002252	EIF3L	Eukaryotic translation initiation factor 3 subunit L
XLOC_002253	eif3l	Eukaryotic translation initiation factor 3 subunit L
XLOC_002256	EaffTmpM012429	Cholinesterase 2
XLOC_002267	pip	Probable proline iminopeptidase
XLOC_002268	EaffTmpA012442	hypothetical protein
XLOC_002271	EaffTmpM012450	Venom allergen 5
XLOC_002272	GLIPR1L1	Venom allergen 3

XLOC_002274	EaffTmpM012452	Glucose-dependent insulinotropic receptor, partial
XLOC_002297	EaffTmpM012423	Genome sequencing data, contig C277
XLOC_002298	EaffTmpM012424	hypothetical protein
XLOC_002301	SLC29A3	Embryonic protein DC-8, partial
XLOC_002308	RPGR	hypothetical protein
XLOC_002310	GSN	Gelsolin
XLOC_002315	EaffTmpA012453	CRISP/Allergen/PR-1
XLOC_002319	EaffTmpA012458	sin3a-associated protein sap130, putative
XLOC_002328	mec 2	Band 7 protein AGAP004871
XLOC_002329	AGAP005782	ATPase ASNA1 homolog
XLOC_002396	FABP9	Fatty acid-binding protein 9
XLOC_002801	EaffTmpM003472	Protein Bm3600, isoform d
XLOC_003587	Xrcc1	DNA repair protein XRCC1
XLOC_004138	EaffTmpA014548	GL12640
XLOC_004384	SLC6A5	Sodium- and chloride-dependent glycine transporter 2
XLOC_004465	EMC1	S phase cyclin A-associated protein in the endoplasmic reticulum
XLOC_004668	Scarb2	Lysosome membrane protein 2
XLOC_005775	AK	Arginine kinase
XLOC_005776	EaffTmpM014180	Protein lava lamp
XLOC_006252	CD109	CD109 antigen
XLOC_006409	HMCN1	Hemicentin-1
XLOC_006418	GABRD	Gamma-aminobutyric acid receptor subunit delta
XLOC_006983	MRPS10	28S ribosomal protein S10, mitochondrial
XLOC_006984	CG8230	Dymeclin
XLOC_006990	Sb	Serine proteinase stubble
XLOC_006991	TASP1	Threonine aspartase 1
XLOC_006998	EaffTmpM016715	conserved Plasmodium protein, unknown function
XLOC_007000	mec 2	Mechanosensory protein 2
XLOC_007001	sto 2	Mechanosensory protein 2
XLOC_007003	Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit
XLOC_007006	stard3	StAR-related lipid transfer protein 3
XLOC_007011	Pcbp3	Poly(rC)-binding protein 3
XLOC_007012	PCBP2	poly(rC)-binding protein 3

XLOC_007016	eipr1	Protein TSSC1
XLOC_007017	Nle1	Notchless protein homolog 1
XLOC_007020	Oxct1	Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial
XLOC_007021	EaffTmpM016693	hypothetical protein Phum PHUM334420
XLOC_007027	Sb	Transmembrane protease serine 12
XLOC_007028	EaffTmpM016703	Serine proteinase stubble
XLOC_007030	GCDH	Glutaryl-CoA dehydrogenase, mitochondrial
XLOC_007031	creld2	Cysteine-rich with EGF-like domain protein 2-B
XLOC_007035	stard3	StAR-related lipid transfer protein 3
XLOC_007037	ARA1	Acetylcholine receptor subunit alpha-like
XLOC_007039	EaffTmpS016728	spore coat protein
XLOC_007040	unc 52	Pikachurin
XLOC_007041	CLEC10A	C-type lectin domain family 10 member A
XLOC_007230	Stard3	StAR-related lipid transfer protein 3
XLOC_007231	stard3	StAR-related lipid transfer protein 3
XLOC_007602	EaffTmpM015504	PREDICTED: kelch-like protein diablo-like
XLOC_007649	Tmprss6	Transmembrane protease serine 6
XLOC_007778	EaffTmpS004307	GH25020
XLOC_007780	Nach-PPK28	Pickpocket protein 28
XLOC_008062	Tret1	Facilitated trehalose transporter Tret1
XLOC_008086	Tret1 2	Facilitated trehalose transporter Tret1-2 homolog
XLOC_008087	Tret1	Facilitated trehalose transporter Tret1
XLOC_009137	GABRR2	Gamma-aminobutyric acid receptor subunit rho-2
XLOC_009164	TTC21B	Tetratricopeptide repeat protein 21B
XLOC_009287	pxt	Chorion peroxidase
XLOC_009586	ccdc85c	Coiled-coil domain-containing protein 85C
XLOC_009588	Dmel Tmp pnn	GA21034
XLOC_009967	ds	Protein dachsous
XLOC_009995	Pnpla2	Patatin-like phospholipase domain-containing protein 2
XLOC_010055	Eif4g3	Eukaryotic translation initiation factor 4 gamma 3
XLOC_010734	EaffTmpM016664	uncharacterized protein LOC319719
XLOC_010995	Ext2	Exostosin-2
XLOC_011027	Fzd1	Frizzled-1

XLOC_011165	PKHD1L1	Fibrocystin-L
XLOC_011203	lilli	AF4/FMR2 family member 4
XLOC_011301	RDH12	Retinol dehydrogenase 13
XLOC_011311	Prss41	Serine protease 41
XLOC_011376	TRMO	Nef-associated protein 1
XLOC_011377	CLS	Probable cardiolipin synthase (CMP-forming)
XLOC_011379	sucg 1	Probable succinyl-CoA ligase
XLOC_011381	rrp45	Exosome complex component rrp45
XLOC_011382	EaffTmpM005151	DNA polymerase III polC-type
XLOC_011415	EaffTmpM005193	serine protease
XLOC_011416	AQP3	Aquaporin-3
XLOC_012278	pip	Heparan sulfate 2-O-sulfotransferase pipe
XLOC_012627	EaffTmpA018418	squash family serine protease inhibitor
XLOC_012629	EaffTmpM018419	Squash family serine protease inhibitor
XLOC_013168	EaffTmpM019162	AGAP004007-PA
XLOC_013618	Chia	Acidic mammalian chitinase
XLOC_013628	eIF1	Protein translation factor SUI1 homolog
XLOC_013633	Phf19	PHD finger protein 19
XLOC_013655	RP1L1	Retinitis pigmentosa 1-like 1 protein
XLOC_013662	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4
XLOC_013665	ATP5L	ATP synthase subunit g, mitochondrial
XLOC_013666	MRPL43	39S ribosomal protein L43, mitochondrial
XLOC_013721	Slc18b1	multidrug transporter
XLOC_013726	prrc1	Protein PRRC1
XLOC_013727	prrc1	Protein PRRC1
XLOC_013729	EaffTmpA020123	AF308673 2 cell surface mucin-like protein
XLOC_013740	Slc18b1	MFS-type transporter SLC18B1
XLOC_013742	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013743	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013747	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013748	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013749	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_014371	CSNK1E	Casein kinase I isoform delta

XLOC_014376	STRAP	Serine-threonine kinase receptor-associated protein
XLOC_014380	TRIM33	Transcription intermediary factor 1-beta
XLOC_014383	HAG1	Histone acetyltransferase GCN5
XLOC_014385	Dhod	Dihydroorotate dehydrogenase (quinone), mitochondrial
XLOC_014388	tmem192	Transmembrane protein 192
XLOC_014392	Acsl4	Long-chain-fatty-acid-CoA ligase 4
XLOC_014393	Acsl3	Long-chain-fatty-acid-CoA ligase 3
XLOC_014397	PKC1	Protein kinase C
XLOC_014400	bicc1 b	Protein bicaudal C homolog 1-B
XLOC_014402	ZFAND5	AN1-type zinc finger protein 5
XLOC_014403	Hydr2	Abhydrolase domain-containing protein 2
XLOC_014404	NSFL1C	NSFL1 cofactor p47
XLOC_014405	pen 2	Gamma-secretase subunit pen-2
XLOC_014406	rnf157	RING finger protein 157
XLOC_014407	Rnf157	Probable E3 ubiquitin-protein ligase MGRN1
XLOC_014408	Rnf157	RING finger protein 157
XLOC_014409	VWDE	von Willebrand factor D and EGF domain-containing protein
XLOC_014410	EaffTmpM019006	von Willebrand factor D and EGF domain-containing protein
XLOC_014411	NOTUM	Protein notum homolog
XLOC_014478	ncs 2	Neuronal calcium sensor 2
XLOC_014506	EaffTmpM022651	EF-hand domain-containing protein D1
XLOC_014625	Slc6a1	Sodium- and chloride-dependent GABA transporter 1
XLOC_014626	SLC6A13	Sodium- and chloride-dependent GABA transporter 1
XLOC_014714	Smlt0970	Peptidyl-Asp metalloendopeptidase
XLOC_014880	nrf 6	Nose resistant to fluoxetine protein 6
XLOC_015278	Gyc88E	Soluble guanylate cyclase 88E
XLOC_015973	rdgA	Eye-specific diacylglycerol kinase
XLOC_017266	EaffTmpM000924	Calmodulin
XLOC_017283	cdk4	Cyclin-dependent kinase 4
XLOC_017284	cdk4	Cyclin-dependent kinase 4
XLOC_017288	EaffTmpM000988	Coiled-coil domain-containing protein 50
XLOC_017304	DENND4A	C-myc promoter-binding protein
XLOC_017306	DENND4C	C-myc promoter-binding protein

XLOC_017307	DENND4A	C-myc promoter-binding protein
XLOC_017359	XCC4067	2-keto-3-deoxy-L-fuconate dehydrogenase
XLOC_017360	Cele Tmp c15a11.4	Protein Bm3600, isoform d
XLOC_017361	EaffTmpM001118	amine oxidase-like protein
XLOC_017441	PNPT1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017442	Vps39	Vam6/Vps39-like protein
XLOC_017450	marf1	Meiosis arrest female protein 1 homolog
XLOC_017451	marf1	Meiosis arrest female protein 1 homolog
XLOC_017453	EHF	ETS homologous factor
XLOC_017474	CML6	Putative calmodulin-like protein 6
XLOC_017483	PDF	Transforming protein Qin
XLOC_017487	EaffTmpM000981	PREDICTED: TCF3 fusion partner-like isoform X2
XLOC_017488	rmdn2	Regulator of microtubule dynamics protein 2
XLOC_017492	LRRC40	Leucine-rich repeat-containing protein 40
XLOC_017494	fbxl15	F-box/LRR-repeat protein 15
XLOC_017566	Shab	Potassium voltage-gated channel protein Shab
XLOC_017572	RHOBTB2	Rho-related BTB domain-containing protein 2
XLOC_017574	RHOBTB1	Rho-related BTB domain-containing protein 1
XLOC_018230	Chia	Acidic mammalian chitinase
XLOC_018666	EaffTmpM006183	hypothetical protein GUITHDRAFT 82324
XLOC_018791	SRSF2	Serine/arginine-rich splicing factor 2
XLOC_018869	EaffTmpM020559	C-type lectin
XLOC_019039	Plg	Plasminogen
XLOC_019205	CTRL	Chymotrypsin-like protease CTRL-1
XLOC_019363	LIN28A	Protein lin-28 homolog A
XLOC_019365	Hdc	Histidine decarboxylase
XLOC_019371	MKKS	McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin
XLOC_019375	EaffTmpM024225	PREDICTED: uncharacterized protein LOC764724
XLOC_019376	EaffTmpM024229	Vascular endothelial growth factor A-A
XLOC_019380	HPGD	15-hydroxyprostaglandin dehydrogenase
XLOC_019385	Dmel Tmp cg13124	MIF4G domain-containing protein-B
XLOC_019787	AMT-6-1	Ammonia Transporter, paralog 6-1
XLOC_019788	AMT-6	Ammonia Transporter, paralog 6

XLOC_019936	EaffTmpM023760	Ovarian abundant message protein, partial
XLOC_019949	KIF28P	Kinesin-like protein KIF28P
XLOC_019951	Kif28p	Kinesin-like protein KIF28P
XLOC_019953	NPEPPS	Aminopeptidase M1-C
XLOC_020410	MYO18A	Unconventional myosin-XVIIIa
XLOC_020602	Pxn	Peroxidasin
XLOC_020603	pxn 1	Peroxidasin homolog
XLOC_020818	pxdn	Peroxidasin
XLOC_023073	trp 1	Transient-receptor-potential-like protein
XLOC_024209	Fubp1	Far upstream element-binding protein 1
XLOC_024322	EaffTmpM006382	Sodium channel protein Nach
XLOC_024623	EaffTmpM025844	Mast cell protease 6 precursor, putative
XLOC_024626	Gie	ADP-ribosylation factor-like protein 8
XLOC_024627	Dhx8	ATP-dependent RNA helicase DHX8
XLOC_024636	slc5a9	Sodium/glucose cotransporter 4
XLOC_024638	CLPP	ATP-dependent Clp protease proteolytic subunit, mitochondrial
XLOC_024856	AGMO	Alkylglycerol monooxygenase
XLOC_024857	Slc13a3	Solute carrier family 13 member 3
XLOC_024863	EaffTmpA025717	Breakpoint cluster region protein
XLOC_024868	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_024869	Slc13a2	Solute carrier family 13 member 2
XLOC_024870	nac 1	Sodium-dependent low-affinity dicarboxylate transporter 1
XLOC_024873	5 HT1B	5-hydroxytryptamine receptor 2B
XLOC_025061	EaffTmpM009103	thioredoxin domain-containing protein, partial
XLOC_025072	PLXNA2	Plexin-A2
XLOC_025073	Plxna4	Plexin-A4
XLOC_025074	PLXNA4	Plexin-A4
XLOC_025084	Lsm2	Thioredoxin-like protein 4A
XLOC_025094	EaffTmpM009149	PREDICTED: uncharacterized protein LOC102634126 isoform X2
XLOC_025099	VWA5A	von Willebrand factor A domain-containing protein 5A
XLOC_025102	Dmel Tmp cg15020	GI12727
XLOC_025115	PEG3	Paternally-expressed gene 3 protein
XLOC_025116	EaffTmpM009100	predicted protein

XLOC_025121	eif3l	Eukaryotic translation initiation factor 3 subunit L
XLOC_025124	VhaAC39 1	V-type proton ATPase subunit d 1
XLOC_025128	apex1	DNA-(apurinic or apyrimidinic site) lyase
XLOC_025142	SLC2A13	sugar transporter
XLOC_025143	SLC2A13	Proton myo-inositol cotransporter
XLOC_025144	Gpdh1	Glycerol-3-phosphate dehydrogenase
XLOC_025146	qvr	Protein quiver
XLOC_025147	kay	Transcription factor kayak
XLOC_025875	ClC a	Chloride channel protein 2
XLOC_025975	PRSS27	Trypsin-1
XLOC_026235	Mhc	Myosin heavy chain, muscle
XLOC_026377	EaffTmpM025593	DNA GyrAse a-subunit
XLOC_026873	EaffTmpM026295	AGAP004872-PA
XLOC_027951	EaffTmpM007749	Excitatory amino acid transporter 1
XLOC_028117	NCAN	Neurocan core protein
XLOC_029225	TFAP4	Transcription factor AP-4
XLOC_029227	EaffTmpM009278	Transcription factor AP-4
XLOC_029242	EaffTmpM009308	Fatty acid-binding protein
XLOC_029660	eng1a	Homeobox protein engrailed-1a
XLOC_029662	en2 b	Homeobox protein engrailed-2-B
XLOC_029663	EaffTmpM028237	F-box/WD repeat-containing protein 1A
XLOC_029665	eng1a	Homeobox protein engrailed-1a
XLOC_029849	bag	IgA FC receptor
XLOC_029916	EaffTmpM028024	Pogo transposable element with ZNF, partial
XLOC_030148	EaffTmpM009811	C-type lectin 5 precursor
XLOC_030209	EaffTmpM027495	Sperm acrosomal protein FSA-ACR.1
XLOC_030440	Nek8	Serine/threonine-protein kinase Nek8
XLOC_030441	nek3	Serine/threonine-protein kinase Nek8
XLOC_030443	EaffTmpM002282	predicted protein
XLOC_030444	F13E6.1	Uncharacterized protein F13E6.1
XLOC_030450	Diras1	GTP-binding protein Di-Ras1
XLOC_030452	PLA2G4A	Cytosolic phospholipase A2
XLOC_030453	Hsap Tmp aoah	Acyloxyacyl hydrolase

XLOC_030458	hddc3	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase MESH1
XLOC_030461	Invadolysin	Leishmanolysin-like peptidase
XLOC_030464	EaffTmpM002323	dynein heavy chain
XLOC_030540	Gxylt2	Glucoside xylosyltransferase 2
XLOC_030558	ACT1	Actin-1
XLOC_030559	Wdr54	WD repeat-containing protein 54
XLOC_030562	Fas1	Fasciclin-1
XLOC_030564	FAS1	Fasciclin-1
XLOC_030565	FAS1	Fasciclin-1
XLOC_030566	Slc25a46	Solute carrier family 25 member 46
XLOC_030570	tdh	L-threonine 3-dehydrogenase
XLOC_030573	CASKIN1	Caskin-1
XLOC_030577	EaffTmpS002307	predicted protein
XLOC_030578	CEP104	Centrosomal protein of 104 kDa
XLOC_030993	NKCC-frag	Na+,K+,2Cl- Cotransporter, fragment
XLOC_031141	PFD1115c	Uncharacterized protein PFD1115c
XLOC_032621	NHA-7	Na+/H+ Antiporter, paralog 7
XLOC_032683	NHA-5	Na+/H+ Antiporter, paralog 5
XLOC_032684	NHA-4	Na+/H+ Antiporter, paralog 4
XLOC_033548	NDUFS3	NADH dehydrogenase
XLOC_033549	NDUFS3	NADH dehydrogenase
XLOC_033551	DDX3X	ATP-dependent RNA helicase DDX3X
XLOC_033922	SON	Protein SON
XLOC_034978	EaffTmpM001985	Tropomyosin
XLOC_035021	Ptp10D	Tyrosine-protein phosphatase 10D
XLOC_035022	Ptp10D	Tyrosine-protein phosphatase 10D
XLOC_035139	NELL1	Protein kinase C-binding protein NELL2
XLOC_035851	NKA-b-5	Na+/K+-ATPase, subunit beta, paralog 5
XLOC_035874	RpLP0	60S acidic ribosomal protein P2
XLOC_035875	PDCD11	Protein RRP5 homolog
XLOC_035876	Pdcd11	Protein RRP5 homolog
XLOC_035877	EaffTmpM011350	Kallikrein-5 precursor
XLOC_035878	KLKB1	Plasma kallikrein

XLOC_035880	EaffTmpM011353	hypothetical protein
XLOC_035882	FN1	PREDICTED: receptor-type tyrosine-protein phosphatase H
XLOC_035966	GPX4	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial
XLOC_035967	TNPO1	Transportin-1
XLOC_035969	Tnpo2	Transportin-2
XLOC_035970	ZNF585A	Zinc finger protein 585A
XLOC_035974	EaffTmpM013740	hypothetical protein KGM 13604
XLOC_035975	MMD2	Monocyte to macrophage differentiation factor
XLOC_036002	EaffTmpM013721	GF21553
XLOC_036007	EaffTmpM013732	PREDICTED: uncharacterized protein DKFZp434B061-like
XLOC_036008	zgc:112255	Uncharacterized protein C1orf50 homolog
XLOC_036011	qvr	Protein quiver
XLOC_036012	EaffTmpM013742	Pupal cuticle protein G1A, putative
XLOC_036013	bgm	Very long-chain-fatty-acid-CoA ligase bubblegum
XLOC_036020	NELF B	Negative elongation factor B
XLOC_036035	Fbln7	c4b-binding protein beta chain, putative
XLOC_036520	Hsp67Ba	Heat shock protein 67B1
XLOC_036961	EaffTmpM010045	hypothetical protein, partial
XLOC_037257	Xdh	Xanthine dehydrogenase
XLOC_037547	SSPO	SCO-spondin

Table S2. Candidate genes under selection in Montmagny or L'Isle Verte, the ancestral saline habitats, detected with F_{ST_MaxSNP}.

Gene ID	**Gene Symbol**	**Description**
XLOC_000267	SLC6A6	Sodium- and chloride-dependent glycine transporter 2
XLOC_000808	EaffTmpM029332	luciferase
XLOC_000820	Glra3	Glycine receptor subunit alpha-3
XLOC_001036	APOD	Apolipoprotein D
XLOC_001043	EaffTmpM012603	hypothetical protein
XLOC_001052	EaffTmpM012607	Serine-aspartate repeat-containing protein I

XLOC_001055	CA-14	Carbonic Anhydrase, paralog 14
XLOC_001062	Acad10	Acyl-CoA dehydrogenase family member 10
XLOC_001063	Tdrd3	Tudor domain-containing protein 3
XLOC_001068	SFXN1	Sideroflexin-1
XLOC_001069	HtrA2	Serine protease HTRA2, mitochondrial
XLOC_001078	EaffTmpS012640	Hypothetical protein CBG24990
XLOC_001080	TNS	Tensin
XLOC_001083	unc45b	Ribonuclease P protein subunit rpr2
XLOC_001095	MYLK	Myosin light chain kinase, smooth muscle
XLOC_001100	add	Adenosine deaminase
XLOC_001101	THAP9	DNA transposase THAP9
XLOC_001105	EaffTmpM012611	RNA-binding protein 12B
XLOC_001111	RAB15	Enamelin
XLOC_001115	VDE1	Violaxanthin de-epoxidase, chloroplastic
XLOC_001120	YAP1	Transcriptional coactivator YAP1
XLOC_001129	Sumo3	Small ubiquitin-related modifier 3
XLOC_001132	NAXD	ATP-dependent (S)-NAD(P)H-hydrate dehydratase
XLOC_001157	Lar	Tyrosine-protein phosphatase Lar
XLOC_001262	EaffTmpM012898	AGAP004574-PA
XLOC_002256	EaffTmpM012429	Cholinesterase 2
XLOC_002264	EaffTmpM012435	hypothetical protein TcasGA2 TC001444
XLOC_002267	pip	Probable proline iminopeptidase
XLOC_002268	EaffTmpA012442	hypothetical protein
XLOC_002271	EaffTmpM012450	Venom allergen 5
XLOC_002272	GLIPR1L1	Venom allergen 3
XLOC_002274	EaffTmpM012452	Glucose-dependent insulinotropic receptor, partial
XLOC_002297	EaffTmpM012423	Genome sequencing data, contig C277
XLOC_002298	EaffTmpM012424	hypothetical protein
XLOC_002301	SLC29A3	Embryonic protein DC-8, partial
XLOC_002308	RPGR	hypothetical protein
XLOC_002309	Gas1	Growth arrest-specific protein 1
XLOC_002315	EaffTmpA012453	CRISP/Allergen/PR-1
XLOC_002316	EaffTmpS012454	Venom allergen 5

XLOC_002319	EaffTmpA012458	sin3a-associated protein sap130, putative
XLOC_002328	mec 2	Band 7 protein AGAP004871
XLOC_002329	AGAP005782	ATPase ASNA1 homolog
XLOC_002440	CHD1	Chromodomain-helicase-DNA-binding protein 1
XLOC_002674	Pisd	Phosphatidylserine decarboxylase proenzyme
XLOC_002741	TLL1	Tolloid-like protein 1
XLOC_002742	EaffTmpM011965	hypothetical protein NEMVEDRAFT v1g47836
XLOC_003110	EaffTmpM008084	Ubiquitin
XLOC_003345	Dmel Tmpw	hypothetical protein L798 12353
XLOC_003390	acsA2	Acetyl-coenzyme A synthetase 2
XLOC_003874	nadK	NAD kinase
XLOC_004053	cpeb1 b	Cytoplasmic polyadenylation element-binding protein 1-B
XLOC_004121	egh	Beta-1,4-mannosyltransferase egh
XLOC_004123	Me3	NADP-dependent malic enzyme
XLOC_004138	EaffTmpA014548	GL12640
XLOC_004314	Hlf	Hepatic leukemia factor
XLOC_004331	Znfx1	NFX1-type zinc finger-containing protein 1
XLOC_004384	SLC6A5	Sodium- and chloride-dependent glycine transporter 2
XLOC_004465	EMC1	S phase cyclin A-associated protein in the endoplasmic reticulum
XLOC_004637	EaffTmpM015926	hypothetical protein LOTGIDRAFT 166205
XLOC_004669	Cd36	Platelet glycoprotein 4
XLOC_004735	Sh	Potassium voltage-gated channel protein Shaker
XLOC_004822	sno1	Senecionine N-oxygenase
XLOC_004868	EaffTmpM013603	Cubilin, partial
XLOC_005093	Mhc	Myosin heavy chain, muscle
XLOC_005147	Iswi	Chromatin-remodeling complex ATPase chain Iswi
XLOC_005427	41333	E3 ubiquitin-protein ligase MARCH8
XLOC_005433	At3g05155	Facilitated trehalose transporter Tret1
XLOC_005775	AK	Arginine kinase
XLOC_005776	EaffTmpM014180	Protein lava lamp
XLOC_005793	hira	28S ribosomal protein S29, mitochondrial
XLOC_006139	Ca P60A	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type
XLOC_006169	QSOX1	Sulfhydryl oxidase 1

XLOC_006221	EaffTmpA013865	Lamin-C
XLOC_006251	CYP3A24	Cytochrome P450 3A16
XLOC_006252	CD109	CD109 antigen
XLOC_006387	EaffTmpA013269	calmin-like protein
XLOC_006409	HMCN1	Hemicentin-1
XLOC_006418	GABRD	Gamma-aminobutyric acid receptor subunit delta
XLOC_006578	EaffTmpM016437	northern shrimp nuclease
XLOC_006983	MRPS10	28S ribosomal protein S10, mitochondrial
XLOC_006984	CG8230	Dymeclin
XLOC_006990	Sb	Serine proteinase stubble
XLOC_006991	TASP1	Threonine aspartase 1
XLOC_006994	SPBPJ4664.02	predicted protein
XLOC_007000	mec 2	Mechanosensory protein 2
XLOC_007001	sto 2	Mechanosensory protein 2
XLOC_007003	Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit
XLOC_007006	stard3	StAR-related lipid transfer protein 3
XLOC_007011	Pcbp3	Poly(rC)-binding protein 3
XLOC_007012	PCBP2	poly(rC)-binding protein 3
XLOC_007016	eipr1	Protein TSSC1
XLOC_007017	Nle1	Notchless protein homolog 1
XLOC_007020	Oxct1	Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial
XLOC_007021	EaffTmpM016693	hypothetical protein Phum PHUM334420
XLOC_007027	Sb	Transmembrane protease serine 12
XLOC_007028	EaffTmpM016703	Serine proteinase stubble
XLOC_007030	GCDH	Glutaryl-CoA dehydrogenase, mitochondrial
XLOC_007031	creld2	Cysteine-rich with EGF-like domain protein 2-B
XLOC_007035	stard3	StAR-related lipid transfer protein 3
XLOC_007040	unc 52	Pikachurin
XLOC_007041	CLEC10A	C-type lectin domain family 10 member A
XLOC_007089	EaffTmpM015272	GH18770
XLOC_007139	ptchd1	Patched domain-containing protein 3
XLOC_007225	EaffTmpM017864	hypothetical protein
XLOC_007230	Stard3	StAR-related lipid transfer protein 3

XLOC_007231	stard3	StAR-related lipid transfer protein 3
XLOC_007233	IQSEC1	IQ motif and SEC7 domain-containing protein 1
XLOC_007566	HNRNPH2	Heterogeneous nuclear ribonucleoprotein H2
XLOC_007595	Grik2	Glutamate receptor ionotropic, kainate 2
XLOC_007602	EaffTmpM015504	PREDICTED: kelch-like protein diablo-like
XLOC_007749	Best3	Bestrophin-3
XLOC_007778	EaffTmpS004307	GH25020
XLOC_007827	Svep1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_008050	Ttpal	Retinaldehyde-binding protein 1-like protein 1
XLOC_008062	Tret1	Facilitated trehalose transporter Tret1
XLOC_008074	EaffTmpA018588	FMRF-amide neuropeptides
XLOC_008086	Tret1 2	Facilitated trehalose transporter Tret1-2 homolog
XLOC_008087	Tret1	Facilitated trehalose transporter Tret1
XLOC_008271	ACOX3	Peroxisomal acyl-coenzyme A oxidase 3
XLOC_008489	SQOR	Sulfide:quinone oxidoreductase, mitochondrial
XLOC_008492	NKA-a-5	Na+/K+-ATPase, subunit alpha, paralog 5
XLOC_008526	KAT5	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial
XLOC_008527	Pdk	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 3, mitochondrial
XLOC_008581	EaffTmpM004812	Neuropilin-1, partial
XLOC_008644	FR	FMRFamide receptor
XLOC_009137	GABRR2	Gamma-aminobutyric acid receptor subunit rho-2
XLOC_009164	TTC21B	Tetratricopeptide repeat protein 21B
XLOC_009586	ccdc85c	Coiled-coil domain-containing protein 85C
XLOC_009621	HCG22	Protein PBMUCL2
XLOC_009958	RASSF9	Ras association domain-containing protein 9
XLOC_009967	ds	Protein dachsous
XLOC_009995	Pnpla2	Patatin-like phospholipase domain-containing protein 2
XLOC_010016	EaffTmpM018020	Tubulin beta chain
XLOC_010033	Ets98B	DNA-binding protein D-ETS-4
XLOC_010034	Spdef	SAM pointed domain-containing Ets transcription factor
XLOC_010733	HR38	Probable nuclear hormone receptor HR38
XLOC_010734	EaffTmpM016664	uncharacterized protein LOC319719

XLOC_010890	EaffTmpM019236	cupin
XLOC_011046	FAM135A	Protein FAM135A
XLOC_011047	FAM135A	Protein FAM135A
XLOC_011113	lig	Protein lingerer
XLOC_011165	PKHD1L1	Fibrocystin-L
XLOC_011203	lilli	AF4/FMR2 family member 4
XLOC_011311	Prss41	Serine protease 41
XLOC_011376	TRMO	Nef-associated protein 1
XLOC_011377	CLS	Probable cardiolipin synthase (CMP-forming)
XLOC_011739	Slc2a1	Solute carrier family 2, facilitated glucose transporter member 1
XLOC_011916	Inx3	Innexin inx2
XLOC_012258	Pli	Protein pellino
XLOC_012278	pip	Heparan sulfate 2-O-sulfotransferase pipe
XLOC_012285	PIK3R4	Metastasis-associated protein MTA3
XLOC_012506	bab2	bric a brac-like protein
XLOC_012542	TRI1	Protein TRI1
XLOC_012543	Smox	Spermine oxidase
XLOC_012627	EaffTmpA018418	squash family serine protease inhibitor
XLOC_012629	EaffTmpM018419	Squash family serine protease inhibitor
XLOC_013453	EaffTmpA003821	Astacin
XLOC_013618	Chia	Acidic mammalian chitinase
XLOC_013628	eIF1	Protein translation factor SUI1 homolog
XLOC_013632	DERL1	Derlin-1
XLOC_013662	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4
XLOC_013663	Dmel Tmp cg4365	Probable hydroxyacylglutathione hydrolase
XLOC_013664	hagh	Hydroxyacylglutathione hydrolase, mitochondrial
XLOC_013665	ATP5L	ATP synthase subunit g, mitochondrial
XLOC_013666	MRPL43	39S ribosomal protein L43, mitochondrial
XLOC_013718	ZDHHC14	Probable palmitoyltransferase ZDHHC14
XLOC_013726	prrc1	Protein PRRC1
XLOC_013727	prrc1	Protein PRRC1
XLOC_013729	EaffTmpA020123	AF308673 2 cell surface mucin-like protein
XLOC_013735	EaffTmpM020092	Low-density lipoprotein receptor class A domain-containing protein 3

XLOC_013736	EaffTmpM020094	Armadillo type fold
XLOC_013737	EaffTmpS020095	Armadillo type fold
XLOC_013740	Slc18b1	MFS-type transporter SLC18B1
XLOC_013742	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013743	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013745	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013747	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013748	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013749	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013757	Ac76E	Adenylate cyclase type 2
XLOC_013888	r	Protein PYR1-3
XLOC_014371	CSNK1E	Casein kinase I isoform delta
XLOC_014380	TRIM33	Transcription intermediary factor 1-beta
XLOC_014383	HAG1	Histone acetyltransferase GCN5
XLOC_014385	Dhod	Dihydroorotate dehydrogenase (quinone), mitochondrial
XLOC_014388	tmem192	Transmembrane protein 192
XLOC_014399	TNXB	Tenascin-X
XLOC_014400	bicc1 b	Protein bicaudal C homolog 1-B
XLOC_014402	ZFAND5	AN1-type zinc finger protein 5
XLOC_014403	Hydr2	Abhydrolase domain-containing protein 2
XLOC_014404	NSFL1C	NSFL1 cofactor p47
XLOC_014405	pen 2	Gamma-secretase subunit pen-2
XLOC_014406	rnf157	RING finger protein 157
XLOC_014407	Rnf157	Probable E3 ubiquitin-protein ligase MGRN1
XLOC_014408	Rnf157	RING finger protein 157
XLOC_014409	VWDE	von Willebrand factor D and EGF domain-containing protein
XLOC_014410	EaffTmpM019006	von Willebrand factor D and EGF domain-containing protein
XLOC_014411	NOTUM	Protein notum homolog
XLOC_014506	EaffTmpM022651	EF-hand domain-containing protein D1
XLOC_014625	Slc6a1	Sodium- and chloride-dependent GABA transporter 1
XLOC_014626	SLC6A13	Sodium- and chloride-dependent GABA transporter 1
XLOC_014637	NKA-a-2	Na+/K+-ATPase, subunit alpha, paralog 2
XLOC_014714	Smlt0970	Peptidyl-Asp metalloendopeptidase

XLOC_014880	nrf 6	Nose resistant to fluoxetine protein 6
XLOC_015025	PRDX1	chromaffin granule amine transporter
XLOC_015690	Zfp26	Zinc finger protein 22
XLOC_015778	RpL15	60S ribosomal protein L15
XLOC_015825	EaffTmpM021144	unnamed protein product, partial
XLOC_015973	rdgA	Eye-specific diacylglycerol kinase
XLOC_015983	Sb	Serine proteinase stubble
XLOC_016203	CG7280	Probable sulfite oxidase, mitochondrial
XLOC_016270	Ppp1r12a	Protein phosphatase 1 regulatory subunit 12A
XLOC_016610	ABCC1	Multidrug resistance-associated protein 1
XLOC_016922	KLHDC8B	Kelch domain-containing protein 8A
XLOC_016923	HMGCR	3-hydroxy-3-methylglutaryl-coenzyme A reductase
XLOC_017082	EaffTmpM019891	63 kDa sperm flagellar membrane protein
XLOC_017266	EaffTmpM000924	Calmodulin
XLOC_017282	EaffTmpM000972	Tropomyosin-2
XLOC_017283	cdk4	Cyclin-dependent kinase 4
XLOC_017284	cdk4	Cyclin-dependent kinase 4
XLOC_017285	4CL	Probable 4-coumarate-CoA ligase 3
XLOC_017288	EaffTmpM000988	Coiled-coil domain-containing protein 50
XLOC_017300	EaffTmpM001006	AGAP000593-PA
XLOC_017301	Mlc1	Myosin light chain alkali
XLOC_017302	Rpl23	60S ribosomal protein L23
XLOC_017303	DENND4A	C-myc promoter-binding protein
XLOC_017304	DENND4A	C-myc promoter-binding protein
XLOC_017306	DENND4C	C-myc promoter-binding protein
XLOC_017307	DENND4A	C-myc promoter-binding protein
XLOC_017437	EaffTmpM000928	PREDICTED: ras-related protein Rab-34, isoform NARR, partial
XLOC_017438	PNPT1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017439	Pnpt1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017440	EaffTmpM000931	CG11337, isoform A
XLOC_017441	PNPT1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017442	Vps39	Vam6/Vps39-like protein
XLOC_017451	marf1	Meiosis arrest female protein 1 homolog

XLOC_017453	EHF	ETS homologous factor
XLOC_017470	ADAM9	Glucose-dependent insulinotropic receptor
XLOC_017474	CML6	Putative calmodulin-like protein 6
XLOC_017483	PDF	Transforming protein Qin
XLOC_017485	Foxg1	Forkhead box protein G1
XLOC_017487	EaffTmpM000981	PREDICTED: TCF3 fusion partner-like isoform X2
XLOC_017488	rmdn2	Regulator of microtubule dynamics protein 2
XLOC_017489	rmdn3	Regulator of microtubule dynamics protein 3
XLOC_017490	CUL5	Cullin-5
XLOC_017491	CUL5	Cullin-5
XLOC_017492	LRRC40	Leucine-rich repeat-containing protein 40
XLOC_017494	fbxl15	F-box/LRR-repeat protein 15
XLOC_017498	Mlc1	Myosin light chain alkali
XLOC_017526	Slc6a5	GA10569
XLOC_017571	svr	Carboxypeptidase M
XLOC_017829	EaffTmpM022959	multiple banded antigen
XLOC_017845	C2orf16	Uncharacterized protein C2orf16
XLOC_017880	EaffTmpM020514	PREDICTED: mucin-2-like
XLOC_017881	EaffTmpM020515	PREDICTED: mucin-2-like
XLOC_018230	Chia	Acidic mammalian chitinase
XLOC_018302	PLB1	hypothetical protein PHAVU 007G184300g
XLOC_018512	SLC4A8	Electroneutral sodium bicarbonate exchanger 1
XLOC_018516	Mrc1	Macrophage mannose receptor 1, partial
XLOC_018791	SRSF2	Serine/arginine-rich splicing factor 2
XLOC_019039	Plg	Plasminogen
XLOC_019060	Adgrg4	Probable G-protein coupled receptor 97
XLOC_019152	MYO18A	Unconventional myosin-XVIIIa
XLOC_019153	MYO18A	Unconventional myosin-XVIIIa
XLOC_019205	CTRL	Chymotrypsin-like protease CTRL-1
XLOC_019333	Pxdn	Peroxidasin homolog
XLOC_019365	Hdc	Histidine decarboxylase
XLOC_019371	MKKS	McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin
XLOC_019372	Dync2li1	Cytoplasmic dynein 2 light intermediate chain 1

XLOC_019375	EaffTmpM024225	PREDICTED: uncharacterized protein LOC764724
XLOC_019376	EaffTmpM024229	Vascular endothelial growth factor A-A
XLOC_019380	HPGD	15-hydroxyprostaglandin dehydrogenase
XLOC_019383	ABCB6	ATP-binding cassette sub-family B member 6, mitochondrial
XLOC_019385	Dmel Tmp cg13124	MIF4G domain-containing protein-B
XLOC_019421	Kif13a	Kinesin-like protein KIF13A
XLOC_019453	Pkd1l2	Polycystic kidney disease protein 1-like 2
XLOC_019675	Inx2	Innexin inx2
XLOC_019787	AMT-6-1	Ammonia Transporter, paralog 6-1
XLOC_019936	EaffTmpM023760	Ovarian abundant message protein, partial
XLOC_019949	KIF28P	Kinesin-like protein KIF28P
XLOC_019951	Kif28p	Kinesin-like protein KIF28P
XLOC_019952	ANPEP	Aminopeptidase N
XLOC_020517	ESTA	Bifunctional acetylxylan esterase/xylanase XynS20E
XLOC_020742	FR	FMRFamide receptor
XLOC_020743	EaffTmpM006942	GH11834
XLOC_020818	pxdn	Peroxidasin
XLOC_021259	CDK14	Cyclin-dependent kinase 14
XLOC_021260	Vg	Vitellogenin
XLOC_021410	EaffTmpM021833	PREDICTED: uncharacterized protein LOC102803158, partial
XLOC_021905	EaffTmpM024070	CG14280, isoform A
XLOC_022288	Adcy2	Adenylate cyclase type 2
XLOC_022307	Igfals	Leucine-rich repeat-containing G-protein coupled receptor 5
XLOC_022308	Lrig3	Carboxypeptidase N subunit 2
XLOC_023663	Gria1	Glutamate receptor 1
XLOC_024209	Fubp1	Far upstream element-binding protein 1
XLOC_024330	TRAPPC2	AGAP011344-PA-like protein
XLOC_024375	DNAH3	Dynein heavy chain 3, axonemal
XLOC_024404	FCABP	Flagellar calcium-binding protein
XLOC_024623	EaffTmpM025844	Mast cell protease 6 precursor, putative
XLOC_024626	Gie	ADP-ribosylation factor-like protein 8
XLOC_024627	Dhx8	ATP-dependent RNA helicase DHX8
XLOC_024636	slc5a9	Sodium/glucose cotransporter 4

XLOC_024638	CLPP	ATP-dependent Clp protease proteolytic subunit, mitochondrial
XLOC_024640	Tbrg4	Protein TBRG4
XLOC_024641	ATPsynCf6	ATP synthase-coupling factor 6, mitochondrial
XLOC_024856	AGMO	Alkylglycerol monooxygenase
XLOC_024857	Slc13a3	Solute carrier family 13 member 3
XLOC_024863	EaffTmpA025717	Breakpoint cluster region protein
XLOC_024866	abr	Active breakpoint cluster region-related protein
XLOC_024867	FBXL20	F-box/LRR-repeat protein 20
XLOC_024868	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_024869	Slc13a2	Solute carrier family 13 member 2
XLOC_024870	nac 1	Sodium-dependent low-affinity dicarboxylate transporter 1
XLOC_024873	5 HT1B	5-hydroxytryptamine receptor 2B
XLOC_025072	PLXNA2	Plexin-A2
XLOC_025073	Plxna4	Plexin-A4
XLOC_025074	PLXNA4	Plexin-A4
XLOC_025084	Lsm2	Thioredoxin-like protein 4A
XLOC_025088	EaffTmpM009142	class B secretin-like G-protein coupled receptor GPRmth5, putative
XLOC_025089	mth2	G-protein coupled receptor Mth2
XLOC_025094	EaffTmpM009149	PREDICTED: uncharacterized protein LOC102634126 isoform X2
XLOC_025099	VWA5A	von Willebrand factor A domain-containing protein 5A
XLOC_025102	Dmel Tmp cg15020	GI12727
XLOC_025109	Dmel Tmp cg6592	Chymotrypsin BI
XLOC_025115	PEG3	Paternally-expressed gene 3 protein
XLOC_025116	EaffTmpM009100	predicted protein
XLOC_025121	eif3l	Eukaryotic translation initiation factor 3 subunit L
XLOC_025124	VhaAC39 1	V-type proton ATPase subunit d 1
XLOC_025128	apex1	DNA-(apurinic or apyrimidinic site) lyase
XLOC_025136	EaffTmpM009127	Circumsporozoite protein
XLOC_025142	SLC2A13	sugar transporter
XLOC_025143	SLC2A13	Proton myo-inositol cotransporter
XLOC_025144	Gpdh1	Glycerol-3-phosphate dehydrogenase
XLOC_025146	qvr	Protein quiver
XLOC_025147	kay	Transcription factor kayak

XLOC_025149	DDB G0286969	von Willebrand factor A domain-containing protein DDB G0286969
XLOC_025151	CHRNA10	Neuronal acetylcholine receptor subunit alpha-4
XLOC_025357	ADGRL1	Latrophilin-1
XLOC_025379	NFE2L1	PHIST domain containing protein
XLOC_025584	Lipf	Gastric triacylglycerol lipase
XLOC_025766	Ptpn9	Tyrosine-protein phosphatase non-receptor type 9
XLOC_025875	ClC a	Chloride channel protein 2
XLOC_026235	Mhc	Myosin heavy chain, muscle
XLOC_026284	Ets1	Protein C-ets-1
XLOC_026889	Mgst1	Microsomal glutathione S-transferase 1
XLOC_027440	CA2	Carbonic anhydrase 2
XLOC_027441	Ca9	Carbonic anhydrase 4
XLOC_027933	EaffTmpM007790	hypothetical protein X975 02403, partial
XLOC_028107	EaffTmpM026805	hypothetical protein TcasGA2 TC004858
XLOC_028117	NCAN	Neurocan core protein
XLOC_028173	EaffTmpM026101	PREDICTED: polyhomeotic-proximal chromatin protein-like isoform X1
XLOC_028831	LIMD1	Paxillin homolog 1
XLOC_029053	Chst11	Carbohydrate sulfotransferase 11
XLOC_029059	Tret1	Facilitated trehalose transporter Tret1
XLOC_029064	AGAP005037	Coiled-coil domain-containing protein AGAP005037
XLOC_029420	Dmel Tmp cg10933	GH21638
XLOC_029421	EaffTmpM026627	Trypsin
XLOC_029594	pbo 4	Na(+)/H(+) exchanger protein 7
XLOC_029595	NHE-X-c	Na+/H+ Exchanger, clade X, paralog c
XLOC_029660	eng1a	Homeobox protein engrailed-1a
XLOC_029662	en2 b	Homeobox protein engrailed-2-B
XLOC_029663	EaffTmpM028237	F-box/WD repeat-containing protein 1A
XLOC_029664	ZNF510	Zinc finger protein 510
XLOC_029665	eng1a	Homeobox protein engrailed-1a
XLOC_029918	Pogz	Pogo transposable element with ZNF domain
XLOC_029939	CTDP1	RNA polymerase II subunit A C-terminal domain phosphatase
XLOC_030121	EaffTmpM009851	GF11443

XLOC_030123	EaffTmpA009853	chitin deacetylase 1
XLOC_030148	EaffTmpM009811	C-type lectin 5 precursor
XLOC_030176	MYO5A	Unconventional myosin-Va
XLOC_030436	slmo	Protein slowmo
XLOC_030440	Nek8	Serine/threonine-protein kinase Nek8
XLOC_030441	nek3	Serine/threonine-protein kinase Nek8
XLOC_030443	EaffTmpM002282	predicted protein
XLOC_030444	F13E6.1	Uncharacterized protein F13E6.1
XLOC_030450	Diras1	GTP-binding protein Di-Ras1
XLOC_030452	PLA2G4A	Cytosolic phospholipase A2
XLOC_030453	Hsap Tmp aoah	Acyloxyacyl hydrolase
XLOC_030458	hddc3	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase MESH1
XLOC_030461	Invadolysin	Leishmanolysin-like peptidase
XLOC_030558	ACT1	Actin-1
XLOC_030559	Wdr54	WD repeat-containing protein 54
XLOC_030562	Fas1	Fasciclin-1
XLOC_030564	FAS1	Fasciclin-1
XLOC_030565	FAS1	Fasciclin-1
XLOC_030566	Slc25a46	Solute carrier family 25 member 46
XLOC_030569	eff	Ubiquitin-conjugating enzyme E2-17 kDa
XLOC_030570	tdh	L-threonine 3-dehydrogenase
XLOC_030573	CASKIN1	Caskin-1
XLOC_030577	EaffTmpS002307	predicted protein
XLOC_030578	CEP104	Centrosomal protein of 104 kDa
XLOC_030581	FNDC3A	Fibronectin type-III domain-containing protein 3a
XLOC_031066	RPII	DNA-directed RNA polymerase II subunit RPB1
XLOC_031536	NPC1	Protein patched homolog 1
XLOC_031745	EaffTmpM028439	GG10482
XLOC_031792	TLL1	Tolloid-like protein 1
XLOC_032602	METTL27	Williams-Beuren syndrome chromosomal region 27 protein
XLOC_032621	NHA-7	Na+/H+ Antiporter, paralog 7
XLOC_032684	NHA-4	Na+/H+ Antiporter, paralog 4
XLOC_032854	pde 5	Probable 3',5'-cyclic phosphodiesterase pde-5

XLOC_032868	cac	hypothetical protein BRAFLDRAFT 124569
XLOC_034122	EDEM3	ER degradation-enhancing alpha-mannosidase-like protein 3
XLOC_034197	Scrt2	Transcriptional repressor scratch 1
XLOC_034250	Hpdl	4-hydroxyphenylpyruvate dioxygenase-like protein
XLOC_034978	EaffTmpM001985	Tropomyosin
XLOC_035022	Ptp10D	Tyrosine-protein phosphatase 10D
XLOC_035121	SCYL2	SCY1-like protein 2
XLOC_035139	NELL1	Protein kinase C-binding protein NELL2
XLOC_035160	EaffTmpM002088	putative BR serine/threonine-protein kinase
XLOC_035541	NR5A2	Nuclear receptor subfamily 5 group A member 2
XLOC_035598	Patronin	Patronin
XLOC_035599	DUSP3	Dual specificity protein phosphatase 3
XLOC_035851	NKA-b-5	Na+/K+-ATPase, subunit beta, paralog 5
XLOC_035874	RpLP0	60S acidic ribosomal protein P2
XLOC_035875	PDCD11	Protein RRP5 homolog
XLOC_035876	Pdcd11	Protein RRP5 homolog
XLOC_035877	EaffTmpM011350	Kallikrein-5 precursor
XLOC_035878	KLKB1	Plasma kallikrein
XLOC_035880	EaffTmpM011353	hypothetical protein
XLOC_035882	FN1	PREDICTED: receptor-type tyrosine-protein phosphatase H
XLOC_035969	Tnpo2	Transportin-2
XLOC_035970	ZNF585A	Zinc finger protein 585A
XLOC_035975	MMD2	Monocyte to macrophage differentiation factor
XLOC_035986	SAR1A	GTP-binding protein SAR1a
XLOC_036002	EaffTmpM013721	GF21553
XLOC_036007	EaffTmpM013732	PREDICTED: uncharacterized protein DKFZp434B061-like
XLOC_036008	zgc:112255	Uncharacterized protein C1orf50 homolog
XLOC_036009	fd59A	Fork head domain-containing protein FD3
XLOC_036012	EaffTmpM013742	Pupal cuticle protein G1A, putative
XLOC_036016	EaffTmpM013752	Compound eye opsin BCRH2
XLOC_036019	Dmel Tmp ckd	GH15777
XLOC_036020	NELF B	Negative elongation factor B
XLOC_036024	phnW	2-aminoethylphosphonate-pyruvate transaminase 1

XLOC_036035	Fbln7	c4b-binding protein beta chain, putative
XLOC_036082	sodF	Probable superoxide dismutase
XLOC_036123	bab2	Protein bric-a-brac 2
XLOC_036181	EaffTmpM029053	GD16438
XLOC_036520	Hsp67Ba	Heat shock protein 67B1
XLOC_036753	EaffTmpM001895	hypothetical protein DAPPUDRAFT 224791
XLOC_036756	RBFOX1	RNA binding protein fox-1 homolog 1
XLOC_036921	nog3	Noggin-3
XLOC_037257	Xdh	Xanthine dehydrogenase
XLOC_037432	rsph10b	Radial spoke head 10 homolog B
XLOC_037547	SSPO	SCO-spondin
XLOC_037753	NBC	Na+, HCO3- cotransporter
XLOC_037782	CEP290	Centrosomal protein of 290 kDa

Table S3. Candidate genes under selection detected in Lake Ontario with $PBE_{\_Window}$ against the ancestral saline populations (Montmagny and L'Isle Verte).

Gene ID	**Gene Symbol**	**Description**
XLOC_000340	Xpnpep1	Xaa-Pro aminopeptidase 1
XLOC_000454	LAC	Lachesin
XLOC_001070	mtp 18	Mitochondrial fission process protein 1
XLOC_001095	MYLK	Myosin light chain kinase, smooth muscle
XLOC_001100	add	Adenosine deaminase
XLOC_001120	YAP1	Transcriptional coactivator YAP1
XLOC_001148	trmt10a	tRNA methyltransferase 10 homolog A
XLOC_001150	WDR20	WD repeat-containing protein 20
XLOC_001360	Apeh	Acylamino-acid-releasing enzyme
XLOC_001376	tmem38b a	Trimeric intracellular cation channel type B
XLOC_001418	PGAP2	Post-GPI attachment to proteins factor 2

XLOC_001800	RSU1	Ras suppressor protein 1
XLOC_001802	slc25a30	Kidney mitochondrial carrier protein 1
XLOC_001804	EaffTmpM003427	PREDICTED: uncharacterized protein LOC579487
XLOC_001905	Taf7	Heat shock 70 kDa protein 4L
XLOC_001922	FDX1	Adrenodoxin, mitochondrial
XLOC_002077	Klhl31	CD209 antigen-like protein C
XLOC_002189	EaffTmpM013119	PREDICTED: uncharacterized protein LOC100167407
XLOC_002222	ATP6V0B	Egl nine homolog 3
XLOC_002223	Egln1	Egl nine homolog 1
XLOC_002650	CAT	Catalase
XLOC_002834	Neurl4	Neuralized-like protein 4
XLOC_002930	scrn3	Secernin-3
XLOC_003291	EaffTmpM014660	glycosyltransferase PglE
XLOC_003471	Klf15	Krueppel-like factor 15
XLOC_003702	bath 43	TD and POZ domain-containing protein 3
XLOC_003798	SLC13A2	Solute carrier family 13 member 2
XLOC_003799	SLC13A5	Solute carrier family 13 member 5
XLOC_003844	Uggt1	UDP-glucose:glycoprotein glucosyltransferase 1
XLOC_003917	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_004018	EaffTmpM003173	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta
XLOC_004192	Pdzd2	PDZ domain-containing protein 2
XLOC_004248	Myo1e	Unconventional myosin-Ie
XLOC_004254	NOX5	Calcineurin subunit B type 2
XLOC_004329	ACAC	Acetyl-CoA carboxylase 1
XLOC_004335	EaffTmpM012241	PREDICTED: putative inorganic phosphate cotransporter
XLOC_004429	Adam10	Disintegrin and metalloproteinase domain-containing protein 10
XLOC_004462	CHD6	Chromodomain-helicase-DNA-binding protein 6
XLOC_004464	fus	RNA-binding protein fusilli
XLOC_004465	EMC1	S phase cyclin A-associated protein in the endoplasmic reticulum
XLOC_004930	EaffTmpM013657	Endocuticle structural glycoprotein SgAbd-9
XLOC_004991	aldh8a1	Aldehyde dehydrogenase family 8 member A1
XLOC_005284	Ca alpha1D	Voltage-dependent calcium channel type D subunit alpha-1
XLOC_005517	Rasl11a	conserved hypothetical protein

XLOC_005704	SPAC1F5.02	Protein disulfide-isomerase A5
XLOC_005722	abhd12	Monoacylglycerol lipase ABHD12
XLOC_005765	HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein
XLOC_005766	LAMA1	Laminin subunit alpha-1
XLOC_006245	EaffTmpM013857	hypothetical protein
XLOC_006370	HMCN1	Hemicentin-1
XLOC_006409	HMCN1	Hemicentin-1
XLOC_006662	Dmel Tmp cg42700	conserved hypothetical protein
XLOC_006981	Slc18b1	MFS-type transporter SLC18B1
XLOC_007003	Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit
XLOC_007014	ppl	Glycine cleavage system H protein, mitochondrial
XLOC_007015	SRP72	Signal recognition particle subunit SRP72
XLOC_007037	ARA1	Acetylcholine receptor subunit alpha-like
XLOC_007107	EaffTmpM015245	GL18374
XLOC_007139	ptchd1	Patched domain-containing protein 3
XLOC_007207	EaffTmpM017845	Tubulin alpha-1 chain
XLOC_007293	GAL3ST1	Galactosylceramide sulfotransferase
XLOC_007763	EaffTmpM004273	AGAP006721-PA
XLOC_007829	Vps37b	Vacuolar protein sorting-associated protein 37B
XLOC_008259	SLIT1	Slit homolog 1 protein
XLOC_008380	NPHS1	Nephrin
XLOC_008682	RPII	ABR027Cp, related
XLOC_008917	SORBS2	Sorbin and SH3 domain-containing protein 2
XLOC_009035	pyc 1	Pyruvate carboxylase, mitochondrial
XLOC_009079	flad1	FAD synthase
XLOC_009122	Edg78E	Pupal cuticle protein Edg-78E
XLOC_009390	atg101	Autophagy-related protein 101
XLOC_009411	Rab6	Ras-related protein Rab6
XLOC_009923	Cacna1g	Voltage-dependent T-type calcium channel subunit alpha-1I
XLOC_010016	EaffTmpM018020	Tubulin beta chain
XLOC_010903	spn1	S-antigen protein
XLOC_011136	EaffTmpM018667	cupin

XLOC_012406	Dock2	Dedicator of cytokinesis protein 2
XLOC_012482	rsph1	Radial spoke head 1 homolog
XLOC_012563	nhr 7	Nuclear hormone receptor family member nhr-3
XLOC_012572	unc 89	Muscle M-line assembly protein unc-89
XLOC_012593	EaffTmpM016928	Glutathione S-transferase
XLOC_012731	Fam193a	Protein FAM193A
XLOC_012948	OVCH1	Ovochymase-1
XLOC_013066	Msx2	Homeobox protein MSX-2
XLOC_013165	Ugt	UDP-glucose:glycoprotein glucosyltransferase
XLOC_013291	NPHS1	Kin of IRRE-like protein 2
XLOC_013433	EaffTmpM003787	Tropomyosin
XLOC_013499	Mhc	Myosin heavy chain, muscle
XLOC_013500	Mhc	AF479654 1 Thr-Ser protein
XLOC_013520	Dmel Tmp cg30089	CG30089, isoform B
XLOC_013525	EaffTmpM003818	hypothetical protein NEMVEDRAFT v1g222368
XLOC_013588	STOML2	Stomatin-like protein 2, mitochondrial
XLOC_013613	PRORSD1P	Putative prolyl-tRNA synthetase associated domain-containing protein 1
XLOC_013689	Nf1	Neurofibromin
XLOC_013970	Rbm45	hypothetical protein, partial
XLOC_014140	Cele Tmp k11b4.1	28S ribosomal protein S27, mitochondrial
XLOC_014141	EaffTmpM004578	Mitochondrial 28S ribosomal protein S27
XLOC_014240	EaffTmpS020322	putative peptidase S8/S53 subtilisin kexin sedolisin
XLOC_014388	tmem192	Transmembrane protein 192
XLOC_014409	VWDE	von Willebrand factor D and EGF domain-containing protein
XLOC_014410	EaffTmpM019006	von Willebrand factor D and EGF domain-containing protein
XLOC_014478	ncs 2	Neuronal calcium sensor 2
XLOC_014506	EaffTmpM022651	EF-hand domain-containing protein D1
XLOC_014625	Slc6a1	Sodium- and chloride-dependent GABA transporter 1
XLOC_014637	NKA-a-2	Na+/K+-ATPase, subunit alpha, paralog 2
XLOC_014638	EaffTmpM020004	PREDICTED: uncharacterized protein C18orf63 homolog
XLOC_014641	ZNF273	Transmembrane protein 39A
XLOC_014659	LPIN3	Phosphatidate phosphatase LPIN3

XLOC_014758	C1qtnf4	Complement C1q tumor necrosis factor-related protein 4
XLOC_014840	RpL4	60S ribosomal protein L4
XLOC_014890	allc	Allantoicase
XLOC_015058	fl(2)d	Protein disulfide-isomerase A5
XLOC_015087	NHE2 5	E affinis NHE2 5
XLOC_015156	HD 0322	RutC family protein HD 0322
XLOC_015273	cher	Filamin-B
XLOC_015279	Gyc88E	Soluble guanylate cyclase 88E
XLOC_015385	EaffTmpS021717	putative ionotropic receptor IR25a
XLOC_015786	EaffTmpA021084	Transposon TX1 uncharacterized 149 kDa protein
XLOC_015897	Mkx	Homeobox protein Mohawk
XLOC_016237	Ppfia3	Liprin-alpha-3
XLOC_016239	itsn1	Intersectin-2
XLOC_016298	SLU7	Transcription initiation factor TFIID subunit 6
XLOC_016321	Tctp	Translationally-controlled tumor protein homolog
XLOC_016350	Nckx30C	Sodium/potassium/calcium exchanger Nckx30C
XLOC_016528	Plc21C	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II
XLOC_017447	Dclre1c	Protein artemis
XLOC_017971	Fuca	Putative alpha-L-fucosidase
XLOC_018341	Xxylt1	Xyloside xylosyltransferase 1
XLOC_018495	FR	FMRFamide receptor
XLOC_018498	EaffTmpM023822	Muscle LIM protein Mlp84B
XLOC_018666	EaffTmpM006183	hypothetical protein GUITHDRAFT 82324
XLOC_018745	EaffTmpM006200	hypothetical protein LOTGIDRAFT 89004, partial
XLOC_018857	EaffTmpM022124	Putative urea active transporter 1
XLOC_018891	Prss52	Serine protease 52
XLOC_018940	Scrt1	Transcriptional repressor scratch 1
XLOC_019153	MYO18A	Unconventional myosin-XVIIIa
XLOC_019193	eny2 2	Transcription and mRNA export factor ENY2
XLOC_019272	ball	Nucleosomal histone kinase 1
XLOC_019289	Zfp26	Zinc finger protein 287
XLOC_019372	Dync2li1	Cytoplasmic dynein 2 light intermediate chain 1
XLOC_019497	Aqp3	Aquaporin-3

XLOC_019565	E(z)	Histone-lysine N-methyltransferase EZH2
XLOC_019702	Shab	Shaker-related potassium channel tsha2
XLOC_019786	marf1	Meiosis arrest female protein 1 homolog
XLOC_019871	PLA2G3	Acidic phospholipase A2 PA4
XLOC_019949	KIF28P	Kinesin-like protein KIF28P
XLOC_020345	EaffTmpM016330	Trypsin-1
XLOC_020571	Smyd3	Histone-lysine N-methyltransferase SMYD3
XLOC_020585	APOD	Apolipoprotein D
XLOC_020681	MCF2	Putative neutrophil cytosol factor 1C
XLOC_020867	NKA-a-1	Na+/K+-ATPase, subunit alpha, paralog 1
XLOC_020894	SSR2	Translocon-associated protein subunit beta
XLOC_020937	ARPC3	Actin-related protein 2/3 complex subunit 3
XLOC_021155	Zfp26	Zinc finger protein 142
XLOC_021363	AAEL006169	Lysosomal aspartic protease
XLOC_021378	EaffTmpM022879	Trehalase
XLOC_021426	Plscr2	Phospholipid scramblase 2
XLOC_021574	BRWD1	Bromodomain and WD repeat-containing protein 1
XLOC_021783	Kcnip1	Kv channel-interacting protein 1
XLOC_021828	Arsj	Arylsulfatase J
XLOC_021829	ARSJ	Arylsulfatase J
XLOC_022129	Invadolysin	Leishmanolysin-like peptidase
XLOC_022193	Gpr119	Glucose-dependent insulinotropic receptor
XLOC_022194	Orct	Organic cation transporter protein
XLOC_022196	HPGD	15-hydroxyprostaglandin dehydrogenase
XLOC_022251	hsdl2	Hydroxysteroid dehydrogenase-like protein 2
XLOC_022256	ARMC2	Armadillo repeat-containing protein 2
XLOC_022263	Orct2	Solute carrier family 22 member 8
XLOC_022266	SSNA1	BET1 homolog
XLOC_022470	EaffTmpM024114	tRNA 2-selenouridine synthase
XLOC_022586	MANBA	Beta-mannosidase
XLOC_022972	XYLB	Xylulose kinase
XLOC_023419	eIF3a	F-box protein At5g06550
XLOC_023542	EaffTmpM025085	pro-neuregulin-1, membrane-bound isoform isoform IIa

XLOC_023554	RAX	Retinal homeobox protein Rax
XLOC_023941	EaffTmpM025816	set and mynd domain-containing protein, putative
XLOC_024322	EaffTmpM006382	Sodium channel protein Nach
XLOC_024336	STARD10	PCTP-like protein
XLOC_024517	GH17388	E3 UFM1-protein ligase 1 homolog
XLOC_024868	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_024932	EaffTmpM027026	predicted protein
XLOC_024972	CG43867	Pleckstrin domain-containing family H member 1, partial
XLOC_024987	Sardh	Sarcosine dehydrogenase, mitochondrial
XLOC_025073	Plxna4	Plexin-A4
XLOC_025074	PLXNA4	Plexin-A4
XLOC_025116	EaffTmpM009100	predicted protein
XLOC_025341	Capn11	Calpain-11
XLOC_025342	CAPN8	Calpain-8
XLOC_025355	LanA	Laminin-like protein epi-1
XLOC_025356	EaffTmpM025286	PREDICTED: isovaleryl-CoA dehydrogenase, mitochondrial
XLOC_025357	ADGRL1	Latrophilin-1
XLOC_025358	EaffTmpM025290	Putative uncharacterized protein ENSP00000383309, partial
XLOC_025396	EaffTmpM026192	300 kDa antigen AG231, putative
XLOC_025589	Map3k4	Mitogen-activated protein kinase kinase kinase 4
XLOC_025590	MAP3K4	Mitogen-activated protein kinase kinase kinase 4
XLOC_025712	ZNF425	Zinc finger protein 226
XLOC_025738	SLC24A5	Sodium/potassium/calcium exchanger 5
XLOC_025823	smc3	Structural maintenance of chromosomes protein 3
XLOC_025847	KLF13	Krueppel-like factor 10
XLOC_025867	EaffTmpM009007	predicted protein
XLOC_025905	EaffTmpM008982	Hexuronate transporter
XLOC_026108	SORBS2	Sorbin and SH3 domain-containing protein 2
XLOC_026242	SPBC776.05	Uncharacterized membrane protein C776.05
XLOC_026303	ABTB2	Ankyrin repeat and BTB/POZ domain-containing protein 2
XLOC_026475	FERMT1	Unc-112-related protein
XLOC_026773	EaffTmpM026317	Tropomyosin
XLOC_026784	EaffTmpM026317	glutamic acid-rich protein precursor, putative

XLOC_026929	Gipc1	PDZ domain-containing protein GIPC1
XLOC_027105	RIMBP2	RIMS-binding protein 2
XLOC_027371	X element\ORF2	Probable RNA-directed DNA polymerase from transposon X-element
XLOC_027561	trf3	Tricorn protease-interacting factor F3
XLOC_027562	EaffTmpM007926	aminopeptidase N isoform 1
XLOC_027655	STARD7	StAR-related lipid transfer protein 7, mitochondrial
XLOC_027742	Ash1l	Histone-lysine N-methyltransferase ash1
XLOC_027894	ADK	Adenosine kinase
XLOC_027955	Mmp9	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_028276	sll0108	Putative ammonium transporter sll0108
XLOC_028316	EaffTmpS008751	Venom dipeptidyl peptidase 4
XLOC_028443	EaffTmpA027443	Transposon TX1 uncharacterized 149 kDa protein
XLOC_028455	EaffTmpM025922	Hemocyanin B chain
XLOC_028523	EaffTmpM026424	Putative defense protein Hdd11-like
XLOC_028618	EaffTmpM008853	Alcohol dehydrogenase class-3 chain L
XLOC_028709	ncs1	Calcium-binding protein NCS-1
XLOC_028782	gpaA	Guanine nucleotide-binding protein alpha-1 subunit
XLOC_028971	ZNF668	Zinc finger protein 182
XLOC_028972	Pnkp	Myoneurin
XLOC_029064	AGAP005037	Coiled-coil domain-containing protein AGAP005037
XLOC_029135	PIM3	Serine/threonine-protein kinase pim-3
XLOC_029253	EaffTmpS009329	predicted protein
XLOC_029254	EaffTmpM009330	SPT transcription factor family member
XLOC_029513	glvI	Proton-gated ion channel
XLOC_029598	Htr1d	5-hydroxytryptamine receptor
XLOC_029640	Abcc10	Protein VAC14 homolog
XLOC_030292	Inx3	Innexin inx2
XLOC_030440	Nek8	Serine/threonine-protein kinase Nek8
XLOC_030441	nek3	Serine/threonine-protein kinase Nek8
XLOC_030443	EaffTmpM002282	predicted protein
XLOC_030559	Wdr54	WD repeat-containing protein 54
XLOC_030562	Fas1	Fasciclin-1

XLOC_030570	tdh	L-threonine 3-dehydrogenase
XLOC_030707	pxt	Chorion peroxidase
XLOC_030750	arhgap32	Rho GTPase-activating protein 32
XLOC_030797	otof	Dysferlin
XLOC_030831	Dazap1	DAZ-associated protein 1
XLOC_030859	heatr5b	HEAT repeat-containing protein 5B
XLOC_030861	heatr5a	HEAT repeat-containing protein 5A
XLOC_030879	Usp36	Ubiquitin carboxyl-terminal hydrolase 36
XLOC_030995	NKCC-3	Na+,K+,2Cl- Cotransporter, paralog 3
XLOC_030997	NKCC2	Na+,K+,2Cl- Cotransporter, paralog 2
XLOC_030999	SCAF8	Protein SCAF8
XLOC_031221	ulk3	Serine/threonine-protein kinase ULK3
XLOC_031229	EaffTmpM028846	Elongation factor 1-beta
XLOC_031894	IIV6 235L	type 11 methyltransferase
XLOC_031961	EaffTmpM009411	cuticular protein analogous to peritrophins 3-A2
XLOC_031992	ITPR2	Inositol 1,4,5-trisphosphate receptor type 2
XLOC_032028	Rmnd5a	Protein yippee-like 5
XLOC_032074	GluClalpha	Glutamate-gated chloride channel
XLOC_032137	pou2f1	POU domain, class 2, transcription factor 1
XLOC_032150	EaffTmpM008549	PREDICTED: uncharacterized protein LOC103522330, partial
XLOC_032249	TGFBI	Transforming growth factor-beta-induced protein ig-h3
XLOC_032347	OIT3	Uromodulin
XLOC_032598	Inpp5e	72 kDa inositol polyphosphate 5-phosphatase
XLOC_032604	EaffTmpM008214	hypothetical protein LOTGIDRAFT 143430, partial
XLOC_032620	PCSK2	Neuroendocrine convertase 2
XLOC_032632	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_032668	Tret1 2	Facilitated trehalose transporter Tret1
XLOC_032682	NHA-6	Na+/H+ Antiporter, paralog 6
XLOC_032683	NHA-5	Na+/H+ Antiporter, paralog 5
XLOC_032685	NHA-3	Na+/H+ Antiporter, paralog 3
XLOC_032687	NHA-1-frag	Na+/H+ Antiporter, paralog 1 (fragment)
XLOC_032688	NHA-1	Na+/H+ Antiporter, paralog 1
XLOC_032693	sls	Titin

XLOC_032761	smad2	Mothers against decapentaplegic homolog 3
XLOC_033449	MTR	Methionine synthase
XLOC_033548	NDUFS3	NADH dehydrogenase
XLOC_033549	NDUFS3	NADH dehydrogenase
XLOC_033911	EaffTmpS028717	xylose isomerase
XLOC_033939	LAC	Lachesin
XLOC_033986	EaffTmpM011020	Pecanex-like protein 1
XLOC_034035	PLA2G1B	Phospholipase A2, major isoenzyme
XLOC_034332	nhr 12	Nuclear hormone receptor family member nhr-12
XLOC_034584	rsad2	Radical S-adenosyl methionine domain-containing protein 2
XLOC_034585	ZMYM1	Zinc finger MYM-type protein 1
XLOC_034630	OVCH2	Ovochymase-2
XLOC_034640	SLC18B1	MFS-type transporter SLC18B1
XLOC_034641	EaffTmpM010522	chromaffin granule amine transporter, putative
XLOC_034757	EaffTmpS029207	hypothetical protein DAPPUDRAFT 67487
XLOC_034967	EaffTmpA001971	predicted protein
XLOC_035073	Hn	Protein henna
XLOC_035467	KCNJ18	Inward rectifier potassium channel 18
XLOC_035570	Cyp301a1	Probable cytochrome P450 301a1, mitochondrial
XLOC_035571	CYP12A2	Cytochrome P450 CYP12A2
XLOC_035623	EaffTmpM010889	putative G-protein coupled receptor 112, partial
XLOC_035827	EaffTmpS028872	Protein CBR-CLEC-5
XLOC_035862	TEP1	Telomerase protein component 1
XLOC_035863	TEP1	Telomerase protein component 1
XLOC_035874	RpLP0	60S acidic ribosomal protein P2
XLOC_035875	PDCD11	Protein RRP5 homolog
XLOC_035935	TMEM68	Transmembrane protein 68
XLOC_036016	EaffTmpM013752	Compound eye opsin BCRH2
XLOC_036098	CPO	Carboxypeptidase A5
XLOC_036103	mcm10	Protein MCM10 homolog
XLOC_036139	DHCR24	Delta(24)-sterol reductase
XLOC_036140	DHCR24	Delta(24)-sterol reductase
XLOC_036188	EaffTmpM029230	mucin 2 precursor

XLOC_036449	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_036707	Adarb1	Double-stranded RNA-specific editase Adar
XLOC_036752	EaffTmpM001894	DNA damage-inducible protein
XLOC_036753	EaffTmpM001895	hypothetical protein DAPPUDRAFT 224791
XLOC_036854	AAR2	Protein AAR2 homolog
XLOC_036881	yin	Solute carrier family 15 member 2
XLOC_036940	EaffTmpM009997	GD20657
XLOC_037043	EaffTmpM011765	Tubulin alpha-1D chain
XLOC_037489	CA13	Carbonic anhydrase 13
XLOC_037519	CLNS1A	Putative all-trans-retinol 13,14-reductase
XLOC_037738	myoM	Myosin-M heavy chain
XLOC_037826	SINAT1	Putative E3 ubiquitin-protein ligase SINAT1

Table S4. Candidate genes under selection detected in Lake Ontario with $PBE_{\_MaxSNP}$ against the ancestral saline populations (Montmagny and L'Isle Verte).

Gene ID	**Gene Symbol**	**Description**
XLOC_000412	EaffTmpM000311	PREDICTED: LOW QUALITY PROTEIN: CUB and sushi domain-containing protein 2
XLOC_000592	Rh-2	Rh protein, paralog 2
XLOC_000707	At4g35335	CMP-sialic acid transporter 4
XLOC_001031	NPR2	Atrial natriuretic peptide receptor 2
XLOC_001032	Npr2	Atrial natriuretic peptide receptor 2
XLOC_001036	APOD	Apolipoprotein D
XLOC_001038	EaffTmpM012591	hypothetical protein BRAFLDRAFT 86046
XLOC_001039	EaffTmpM012592	JmjC domain-containing protein C2orf60
XLOC_001043	EaffTmpM012603	hypothetical protein
XLOC_001052	EaffTmpM012607	Serine-aspartate repeat-containing protein I
XLOC_001055	CA-14	Carbonic Anhydrase, paralog 14
XLOC_001058	ACAD10	Acyl-CoA dehydrogenase family member 10

XLOC_001068	SFXN1	Sideroflexin-1
XLOC_001069	HtrA2	Serine protease HTRA2, mitochondrial
XLOC_001070	mtp 18	Mitochondrial fission process protein 1
XLOC_001078	EaffTmpS012640	Hypothetical protein CBG24990
XLOC_001083	unc45b	Ribonuclease P protein subunit rpr2
XLOC_001096	ACSBG2	Long-chain-fatty-acid-CoA ligase ACSBG2
XLOC_001097	ACSBG1	Long-chain-fatty-acid-CoA ligase ACSBG1
XLOC_001098	ACSBG2	Long-chain-fatty-acid-CoA ligase ACSBG2
XLOC_001099	acsbg2	Long-chain-fatty-acid-CoA ligase ACSBG2
XLOC_001101	THAP9	DNA transposase THAP9
XLOC_001103	rhbdf1	Inactive rhomboid protein 1
XLOC_001104	rho 5	Inactive rhomboid protein 1
XLOC_001105	EaffTmpM012611	RNA-binding protein 12B
XLOC_001129	Sumo3	Small ubiquitin-related modifier 3
XLOC_001132	NAXD	ATP-dependent (S)-NAD(P)H-hydrate dehydratase
XLOC_001152	XDH	Probable aldehyde oxidase 3
XLOC_001360	Apeh	Acylamino-acid-releasing enzyme
XLOC_001569	EaffTmpM013039	Arrestin homolog
XLOC_001574	alxA	Alternative oxidase, mitochondrial
XLOC_002195	PRY3	Golgi-associated plant pathogenesis-related protein 1
XLOC_002213	EaffTmpM013099	GM18245
XLOC_002256	EaffTmpM012429	Cholinesterase 2
XLOC_002271	EaffTmpM012450	Venom allergen 5
XLOC_002319	EaffTmpA012458	sin3a-associated protein sap130, putative
XLOC_002328	mec 2	Band 7 protein AGAP004871
XLOC_002340	prkra a	Interferon-inducible double-stranded RNA-dependent protein kinase activator A homolog A
XLOC_002443	SPCC663.09c	Uncharacterized oxidoreductase C663.09c
XLOC_002650	CAT	Catalase
XLOC_002750	STAC2	SH3 and cysteine-rich domain-containing protein 3
XLOC_002801	EaffTmpM003472	Protein Bm3600, isoform d
XLOC_002970	MRVI1	Protein MRVI1
XLOC_003043	ZFP3	Zinc finger protein 3 homolog
XLOC_003216	Pxdn	Peroxidasin-like protein

XLOC_003473	S	Protein Star
XLOC_003518	AAEL011789	Probable citrate synthase 2, mitochondrial
XLOC_003545	SLC27A4	luciferase, putative
XLOC_003548	EaffTmpM015090	endochitinase
XLOC_003799	SLC13A5	Solute carrier family 13 member 5
XLOC_003844	Uggt1	UDP-glucose:glycoprotein glucosyltransferase 1
XLOC_004050	POL	Retrovirus-related Pol polyprotein from transposon 412
XLOC_004138	EaffTmpA014548	GL12640
XLOC_004163	OV16	Putative odorant-binding protein A5
XLOC_004345	sgsm3	Small G protein signaling modulator 3
XLOC_004429	Adam10	Disintegrin and metalloproteinase domain-containing protein 10
XLOC_004455	EaffTmpM012062	conserved hypothetical protein
XLOC_004462	CHD6	Chromodomain-helicase-DNA-binding protein 6
XLOC_004464	fus	RNA-binding protein fusilli
XLOC_004465	EMC1	S phase cyclin A-associated protein in the endoplasmic reticulum
XLOC_004690	SLC46A3	AGAP005317-PA-like protein
XLOC_005028	EaffTmpM002455	Protein F32B5.7, isoform b
XLOC_005110	Gabrr1	Gamma-aminobutyric acid receptor subunit rho-1
XLOC_005113	MEST	Mesoderm-specific transcript homolog protein
XLOC_005194	Hnrnpul1	Heterogeneous nuclear ribonucleoprotein U-like protein 1
XLOC_005231	HSD17B4	Peroxisomal multifunctional enzyme type 2
XLOC_005363	EaffTmpM015686	Putative uncharacterized protein FLJ45035, partial
XLOC_005427	41333	E3 ubiquitin-protein ligase MARCH8
XLOC_005516	EaffTmpA015560	Cuticle protein 7
XLOC_005517	Rasl11a	conserved hypothetical protein
XLOC_005580	SLC16A12	Monocarboxylate transporter 12
XLOC_005606	EaffTmpM015580	unnamed protein product
XLOC_005722	abhd12	Monoacylglycerol lipase ABHD12
XLOC_005724	AP3M1	AP-3 complex subunit mu-1
XLOC_005725	EaffTmpA014069	N-acylglucosamine 2-epimerase
XLOC_005785	ASB2	Ankyrin repeat and SOCS box protein 2
XLOC_006418	GABRD	Gamma-aminobutyric acid receptor subunit delta
XLOC_006419	PSMD13	Thioredoxin-like protein 1

XLOC_006434	Tmbim4	Integrator complex subunit 12
XLOC_006557	GlyP	Glycogen phosphorylase
XLOC_006558	EaffTmpM016394	hypothetical protein BRAFLDRAFT 204830
XLOC_006578	EaffTmpM016437	northern shrimp nuclease
XLOC_006594	abcG23	ABC transporter G family member 23
XLOC_006942	SLC8A1	Sodium/calcium exchanger 1
XLOC_006950	Usp48	Ubiquitin carboxyl-terminal hydrolase 48
XLOC_006990	Sb	Serine proteinase stubble
XLOC_006991	TASP1	Threonine aspartase 1
XLOC_007001	sto 2	Mechanosensory protein 2
XLOC_007003	Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit
XLOC_007006	stard3	StAR-related lipid transfer protein 3
XLOC_007021	EaffTmpM016693	hypothetical protein Phum PHUM334420
XLOC_007063	IQCA1	IQ and AAA domain-containing protein 1
XLOC_007182	EaffTmpA015391	hypothetical protein
XLOC_007213	PHR1	Cryptochrome-1
XLOC_007340	Slit3	Slit homolog 1 protein
XLOC_007436	tmem242	GPI mannosyltransferase 1
XLOC_007447	Fam192a	HMG domain-containing protein 4
XLOC_007595	Grik2	Glutamate receptor ionotropic, kainate 2
XLOC_007666	gpat3	Glycerol-3-phosphate acyltransferase 3
XLOC_007746	Far1	Fatty acyl-CoA reductase 2
XLOC_007749	Best3	Bestrophin-3
XLOC_007778	EaffTmpS004307	GH25020
XLOC_007787	EaffTmpM004325	uncharacterized protein LOC691083
XLOC_007804	KIAA1109	Uncharacterized protein KIAA1109
XLOC_008050	Ttpal	Retinaldehyde-binding protein 1-like protein 1
XLOC_008062	Tret1	Facilitated trehalose transporter Tret1
XLOC_008087	Tret1	Facilitated trehalose transporter Tret1
XLOC_008317	A4galt	Lactosylceramide 4-alpha-galactosyltransferase
XLOC_008581	EaffTmpM004812	Neuropilin-1, partial
XLOC_008610	ACY1	Aminoacylase-1
XLOC_008721	EaffTmpM016818	Transcription factor 25

XLOC_008722	TCF25	Transcription factor 25
XLOC_009035	pyc 1	Pyruvate carboxylase, mitochondrial
XLOC_009074	nrf 6	Nose resistant to fluoxetine protein 6
XLOC_009091	Edg78E	Pupal cuticle protein Edg-78E
XLOC_009123	SLC4A10	Sodium-driven chloride bicarbonate exchanger
XLOC_009137	GABRR2	Gamma-aminobutyric acid receptor subunit rho-2
XLOC_009164	TTC21B	Tetratricopeptide repeat protein 21B
XLOC_009287	pxt	Chorion peroxidase
XLOC_009442	surf1	Surfeit locus protein 1
XLOC_009443	dus3l	tRNA-dihydrouridine(47) synthase
XLOC_009453	EaffTmpM003635	AGAP004872-PA
XLOC_009590	Mical	Protein-methionine sulfoxide oxidase Mical
XLOC_009623	EaffTmpM016249	Hypothetical protein CBG24990
XLOC_009742	UOX	PREDICTED: uricase-like
XLOC_009902	selenbp1 a	Selenium-binding protein 1
XLOC_009958	RASSF9	Ras association domain-containing protein 9
XLOC_010016	EaffTmpM018020	Tubulin beta chain
XLOC_010028	Dmel Tmp crok	GK15491
XLOC_010030	EaffTmpM018046	SET and MYND domain-containing protein
XLOC_010084	Col18a1	Collagen alpha-1(XVIII) chain
XLOC_010654	Dmel Tmp serp	AGAP011936-PA
XLOC_010766	Atg4b	Cysteine protease ATG4B
XLOC_011005	EaffTmpA018384	Cerebellin-3
XLOC_011013	SLC22A5	Solute carrier family 22 member 9
XLOC_011071	LAC	hypothetical protein BRAFLDRAFT 248733
XLOC_011137	DUOX1	Dual oxidase 1
XLOC_011165	PKHD1L1	Fibrocystin-L
XLOC_011301	RDH12	Retinol dehydrogenase 13
XLOC_011311	Prss41	Serine protease 41
XLOC_011377	CLS	Probable cardiolipin synthase (CMP-forming)
XLOC_011379	sucg 1	Probable succinyl-CoA ligase
XLOC_011400	kmt5b	Histone-lysine N-methyltransferase SUV420H1
XLOC_011415	EaffTmpM005193	serine protease

XLOC_011416	AQP3	Aquaporin-3
XLOC_011433	EaffTmpM005213	Otopetrin-2
XLOC_011739	Slc2a1	Solute carrier family 2, facilitated glucose transporter member 1
XLOC_011948	NHP2	H/ACA ribonucleoprotein complex subunit 2-like protein
XLOC_012257	EaffTmpS017774	Fatty acid-binding protein
XLOC_012281	RpA 70	Replication protein A 70 kDa DNA-binding subunit
XLOC_012638	Dnah5	Dynein heavy chain 5, axonemal
XLOC_012709	Ppp4r4	Serine/threonine-protein phosphatase 4 regulatory subunit 4
XLOC_012710	Ppp4r4	Serine/threonine-protein phosphatase 4 regulatory subunit 4
XLOC_012800	MPP5	MAGUK p55 subfamily member 5-A
XLOC_012899	gpaA	Guanine nucleotide-binding protein alpha-1 subunit
XLOC_012980	B4GALT4	Beta-1,4-galactosyltransferase 4
XLOC_013083	Nek8	Serine/threonine-protein kinase Nek8
XLOC_013191	Pdk1	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 3, mitochondrial
XLOC_013265	CYP2G1	Cytochrome P450 2J6
XLOC_013433	EaffTmpM003787	Tropomyosin
XLOC_013458	EaffTmpM003828	Kinesin-like protein KIF19
XLOC_013536	Dmel Tmp cg6040	hypothetical protein Phum PHUM452850
XLOC_013616	EaffTmpM021015	Sptzle 2-like protein
XLOC_013618	Chia	Acidic mammalian chitinase
XLOC_013674	EaffTmpM019126	hypothetical protein
XLOC_013689	Nf1	Neurofibromin
XLOC_013690	ZNF557	Zinc finger protein 557
XLOC_013691	ZFY	Zinc finger protein draculin
XLOC_013726	prrc1	Protein PRRC1
XLOC_013729	EaffTmpA020123	AF308673 2 cell surface mucin-like protein
XLOC_013742	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013747	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013748	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013749	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013856	Ehmt2	Histone-lysine N-methyltransferase EHMT2
XLOC_013863	spe1	Ornithine decarboxylase
XLOC_013884	Nop17l	Protein kintoun

XLOC_014068	EaffTmpA004586	GF15862
XLOC_014144	Slc20a1	Sodium-dependent phosphate transporter 1
XLOC_014380	TRIM33	Transcription intermediary factor 1-beta
XLOC_014383	HAG1	Histone acetyltransferase GCN5
XLOC_014385	Dhod	Dihydroorotate dehydrogenase (quinone), mitochondrial
XLOC_014387	cnpy4	Protein canopy 4
XLOC_014388	tmem192	Transmembrane protein 192
XLOC_014392	Acsl4	Long-chain-fatty-acid-CoA ligase 4
XLOC_014393	Acsl3	Long-chain-fatty-acid-CoA ligase 3
XLOC_014397	PKC1	Protein kinase C
XLOC_014399	TNXB	Tenascin-X
XLOC_014400	bicc1 b	Protein bicaudal C homolog 1-B
XLOC_014406	rnf157	RING finger protein 157
XLOC_014407	Rnf157	Probable E3 ubiquitin-protein ligase MGRN1
XLOC_014408	Rnf157	RING finger protein 157
XLOC_014410	EaffTmpM019006	von Willebrand factor D and EGF domain-containing protein
XLOC_014506	EaffTmpM022651	EF-hand domain-containing protein D1
XLOC_014625	Slc6a1	Sodium- and chloride-dependent GABA transporter 1
XLOC_014637	NKA-a-2	Na+/K+-ATPase, subunit alpha, paralog 2
XLOC_014638	EaffTmpM020004	PREDICTED: uncharacterized protein C18orf63 homolog
XLOC_014645	Tstd3	Thiosulfate sulfurtransferase/rhodanese-like domain-containing protein 3
XLOC_014700	nhr 48	Nuclear hormone receptor family member nhr-48
XLOC_014712	EaffTmpM020045	COG1292: Choline-glycine betaine transporter (ISS)
XLOC_014746	slc25a42	Mitochondrial coenzyme A transporter SLC25A42
XLOC_014880	nrf 6	Nose resistant to fluoxetine protein 6
XLOC_014890	allc	Allantoicase
XLOC_015018	EaffTmpM017157	hypothetical protein DAPPUDRAFT 328622
XLOC_015087	NHE2 5	E affinis NHE2 5
XLOC_015801	slo	Calcium-activated potassium channel slowpoke
XLOC_015825	EaffTmpM021144	unnamed protein product, partial
XLOC_015897	Mkx	Homeobox protein Mohawk
XLOC_015944	T25B9.9	6-phosphogluconate dehydrogenase, decarboxylating
XLOC_016001	Dmel Tmp cg6870	Cytochrome b5

XLOC_016203	CG7280	Probable sulfite oxidase, mitochondrial
XLOC_016210	Ephx1	Epoxide hydrolase 1
XLOC_016228	Dapk1	Death-associated protein kinase 1
XLOC_016298	SLU7	Transcription initiation factor TFIID subunit 6
XLOC_016320	Lrrc15	Leucine-rich repeat-containing protein 15
XLOC_016321	Tctp	Translationally-controlled tumor protein homolog
XLOC_016329	acr 16	Acetylcholine receptor subunit alpha-type acr-16
XLOC_016479	GABRB3	Gamma-aminobutyric acid receptor subunit beta-3
XLOC_016512	EaffTmpM017815	hypothetical protein TcasGA2 TC002700
XLOC_016843	cut 1	Cuticlin-1
XLOC_016844	KSR2	Kinase suppressor of Ras 2
XLOC_017060	KIF4A	Chromosome-associated kinesin KIF4A
XLOC_017092	Dmel Tmp dh44 r2	Diuretic hormone receptor
XLOC_017245	SVEP1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_017253	IQSEC1	IQ motif and SEC7 domain-containing protein 1
XLOC_017282	EaffTmpM000972	Tropomyosin-2
XLOC_017288	EaffTmpM000988	Coiled-coil domain-containing protein 50
XLOC_017300	EaffTmpM001006	AGAP000593-PA
XLOC_017301	Mlc1	Myosin light chain alkali
XLOC_017307	DENND4A	C-myc promoter-binding protein
XLOC_017308	DENND4B	DENN domain-containing protein 4B
XLOC_017360	Cele Tmp c15a11.4	Protein Bm3600, isoform d
XLOC_017361	EaffTmpM001118	amine oxidase-like protein
XLOC_017379	EaffTmpM001145	Antho-RFamide neuropeptides type 2
XLOC_017380	EaffTmpM001146	Putative uncharacterized protein FLJ45035, partial
XLOC_017423	oxlT	Oxalate:formate antiporter
XLOC_017437	EaffTmpM000928	PREDICTED: ras-related protein Rab-34, isoform NARR, partial
XLOC_017438	PNPT1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017439	Pnpt1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017440	EaffTmpM000931	CG11337, isoform A
XLOC_017441	PNPT1	Polyribonucleotide nucleotidyltransferase 1, mitochondrial
XLOC_017447	Dclre1c	Protein artemis

XLOC_017451	marf1	Meiosis arrest female protein 1 homolog
XLOC_017453	EHF	ETS homologous factor
XLOC_017492	LRRC40	Leucine-rich repeat-containing protein 40
XLOC_017498	Mlc1	Myosin light chain alkali
XLOC_017502	mig 15	Serine/threonine-protein kinase mig-15
XLOC_017503	TNIK	TRAF2 and NCK-interacting protein kinase
XLOC_017566	Shab	Potassium voltage-gated channel protein Shab
XLOC_017571	svr	Carboxypeptidase M
XLOC_017572	RHOBTB2	Rho-related BTB domain-containing protein 2
XLOC_017620	Sgta	Small glutamine-rich tetratricopeptide repeat-containing protein alpha
XLOC_017880	EaffTmpM020514	PREDICTED: mucin-2-like
XLOC_018081	EaffTmpM022043	CD109 antigen
XLOC_018522	PPO3	Hemocyanin F chain
XLOC_018666	EaffTmpM006183	hypothetical protein GUITHDRAFT 82324
XLOC_018710	RRBP1	Mannan-binding lectin serine protease 2
XLOC_018791	SRSF2	Serine/arginine-rich splicing factor 2
XLOC_018810	EaffTmpM006285	STI1-like protein
XLOC_018811	EaffTmpM006286	predicted protein
XLOC_018891	Prss52	Serine protease 52
XLOC_019039	Plg	Plasminogen
XLOC_019100	PPCDC	Acyl-CoA:lysophosphatidylglycerol acyltransferase 1
XLOC_019249	DPP6	Inactive dipeptidyl peptidase 10
XLOC_019272	ball	Nucleosomal histone kinase 1
XLOC_019297	CadN	Neural-cadherin
XLOC_019365	Hdc	Histidine decarboxylase
XLOC_019371	MKKS	McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin
XLOC_019372	Dync2li1	Cytoplasmic dynein 2 light intermediate chain 1
XLOC_019380	HPGD	15-hydroxyprostaglandin dehydrogenase
XLOC_019383	ABCB6	ATP-binding cassette sub-family B member 6, mitochondrial
XLOC_019410	Gem	GTP-binding protein GEM
XLOC_019426	BEST1	Bestrophin-4
XLOC_019427	BEST1	Bestrophin-1
XLOC_019691	lolal	Longitudinals lacking protein-like

XLOC_019783	Mfsd8	Major facilitator superfamily domain-containing protein 8
XLOC_019936	EaffTmpM023760	Ovarian abundant message protein, partial
XLOC_019949	KIF28P	Kinesin-like protein KIF28P
XLOC_019952	ANPEP	Aminopeptidase N
XLOC_020005	Plscr1	Phospholipid scramblase 1
XLOC_020035	EaffTmpM023728	Hexuronate transporter
XLOC_020036	SPNS2	Protein spinster homolog 2
XLOC_020203	CG5098	Transcription factor 20
XLOC_020255	EaffTmpM007243	hypothetical protein CAPTEDRAFT 221435
XLOC_020256	PF14 0175	Protein PF14 0175
XLOC_020269	EaffTmpM007163	GJ16362
XLOC_020270	EaffTmpM007164	PREDICTED: hypothetical protein LOC100645147
XLOC_020345	EaffTmpM016330	Trypsin-1
XLOC_020384	NUDT24	Nudix hydrolase 24, chloroplastic
XLOC_020401	NUDT20	Nudix hydrolase 20, chloroplastic
XLOC_020410	MYO18A	Unconventional myosin-XVIIIa
XLOC_020573	KMT2B	Histone-lysine N-methyltransferase 2B
XLOC_020590	GLRA1	Glycine receptor subunit alpha-1
XLOC_020601	pxdn	Peroxidasin
XLOC_020602	Pxn	Peroxidasin
XLOC_020603	pxn 1	Peroxidasin homolog
XLOC_020688	NT5C2	Cytosolic purine 5'-nucleotidase
XLOC_020742	FR	FMRFamide receptor
XLOC_020743	EaffTmpM006942	GH11834
XLOC_020833	EaffTmpM006969	AGAP005830-PA-like protein
XLOC_020867	NKA-a-1	Na+/K+-ATPase, subunit alpha, paralog 1
XLOC_020892	Prss48	Serine protease 48
XLOC_020894	SSR2	Translocon-associated protein subunit beta
XLOC_021260	Vg	Vitellogenin
XLOC_021278	dnc	PREDICTED: cAMP-specific 3',5'-cyclic phosphodiesterase-like isoform 2
XLOC_021363	AAEL006169	Lysosomal aspartic protease
XLOC_021378	EaffTmpM022879	Trehalase
XLOC_021435	EaffTmpM021841	Collagen alpha-2(IV) chain

XLOC_021622	fzd5	Frizzled-8
XLOC_021758	FAM186A	Protein FAM186A
XLOC_021810	HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1
XLOC_021870	STXBP5	Syntaxin-binding protein 5-like
XLOC_022176	Ada2b	Nuclear migration protein nudC
XLOC_022194	Orct	Organic cation transporter protein
XLOC_022261	Gga3	ADP-ribosylation factor-binding protein GGA3
XLOC_022262	RAB33B	Ras-related protein Rab-33B
XLOC_022263	Orct2	Solute carrier family 22 member 8
XLOC_022288	Adcy2	Adenylate cyclase type 2
XLOC_022409	EaffTmpM024630	PREDICTED: probable rhodanese domain-containing dual specificity protein phosphatase-like
XLOC_022524	fahd2	Fumarylacetoacetate hydrolase domain-containing protein 2
XLOC_022586	MANBA	Beta-mannosidase
XLOC_023171	mical1	Protein-methionine sulfoxide oxidase MICAL2
XLOC_023325	EaffTmpS026087	Protein takeout
XLOC_023350	Srprb	Signal recognition particle receptor subunit beta
XLOC_023352	slc38a7	Putative sodium-coupled neutral amino acid transporter 7
XLOC_023412	Dmel Tmp cg31344	AGAP005178-PA-like protein
XLOC_023494	EaffTmpM025414	hypothetical protein
XLOC_023544	ARRDC2	Arrestin domain-containing protein 2
XLOC_023576	snRNP U1 70K	U1 small nuclear ribonucleoprotein 70 kDa
XLOC_023941	EaffTmpM025816	set and mynd domain-containing protein, putative
XLOC_023944	EaffTmpM025801	Tubulin alpha-1 chain
XLOC_023953	B3GALT5	Beta-1,3-galactosyltransferase 5
XLOC_023974	Grn	Granulins
XLOC_024322	EaffTmpM006382	Sodium channel protein Nach
XLOC_024358	EaffTmpM026027	Pancreatic alpha-amylase
XLOC_024364	SGCG	Delta-sarcoglycan
XLOC_024375	DNAH3	Dynein heavy chain 3, axonemal
XLOC_024594	RpL18	60S ribosomal protein L18
XLOC_024641	ATPsynCf6	ATP synthase-coupling factor 6, mitochondrial
XLOC_024727	KHK	Ketohexokinase

XLOC_024825	Smal 0815	Peptidyl-Asp metalloendopeptidase
XLOC_024856	AGMO	Alkylglycerol monooxygenase
XLOC_024857	Slc13a3	Solute carrier family 13 member 3
XLOC_024863	EaffTmpA025717	Breakpoint cluster region protein
XLOC_024867	FBXL20	F-box/LRR-repeat protein 20
XLOC_024868	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_024869	Slc13a2	Solute carrier family 13 member 2
XLOC_024895	Alk	Leukocyte tyrosine kinase receptor
XLOC_024899	EaffTmpM020389	GL17228
XLOC_024909	Scyl2	SCY1-like protein 2
XLOC_024949	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_024966	snrpd3	Small nuclear ribonucleoprotein Sm D3
XLOC_024972	CG43867	Pleckstrin domain-containing family H member 1, partial
XLOC_024973	CG43867	Uncharacterized protein CG43867
XLOC_025019	FOXO6	Forkhead box protein O6
XLOC_025072	PLXNA2	Plexin-A2
XLOC_025073	Plxna4	Plexin-A4
XLOC_025074	PLXNA4	Plexin-A4
XLOC_025102	Dmel Tmp cg15020	GI12727
XLOC_025115	PEG3	Paternally-expressed gene 3 protein
XLOC_025116	EaffTmpM009100	predicted protein
XLOC_025124	VhaAC39 1	V-type proton ATPase subunit d 1
XLOC_025143	SLC2A13	Proton myo-inositol cotransporter
XLOC_025313	METTL18	Histidine protein methyltransferase 1 homolog
XLOC_025314	Tmem63b	CSC1-like protein 2
XLOC_025355	LanA	Laminin-like protein epi-1
XLOC_025358	EaffTmpM025290	Putative uncharacterized protein ENSP00000383309, partial
XLOC_025473	Dmel Tmp eno	Enolase
XLOC_025589	Map3k4	Mitogen-activated protein kinase kinase kinase 4
XLOC_025590	MAP3K4	Mitogen-activated protein kinase kinase kinase 4
XLOC_025675	Ifi30	Gamma-interferon-inducible lysosomal thiol reductase
XLOC_025712	ZNF425	Zinc finger protein 226
XLOC_025747	Skeletor	Protein Skeletor, isoforms D/E

XLOC_025847	KLF13	Krueppel-like factor 10
XLOC_025891	Ces4a	Carboxylesterase 4A
XLOC_025905	EaffTmpM008982	Hexuronate transporter
XLOC_025933	EaffTmpM009049	Probable nitrile hydratase
XLOC_026124	Gyc88E	Soluble guanylate cyclase 88E
XLOC_026173	Ankrd28	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A
XLOC_026235	Mhc	Myosin heavy chain, muscle
XLOC_026284	Ets1	Protein C-ets-1
XLOC_026380	Klhl7	Kelch-like protein 7
XLOC_026420	ZW	Glucose-6-phosphate 1-dehydrogenase
XLOC_026454	EaffTmpA026771	Leucine-rich repeat neuronal protein 4
XLOC_026889	Mgst1	Microsomal glutathione S-transferase 1
XLOC_026998	ORF99	Putative apoptosis inhibitor ORF99
XLOC_027371	X element\ORF2	Probable RNA-directed DNA polymerase from transposon X-element
XLOC_027423	EaffTmpM027195	Protein TTN-1, isoform d
XLOC_027638	Syx1A	Syntaxin-1A
XLOC_027655	STARD7	StAR-related lipid transfer protein 7, mitochondrial
XLOC_027742	Ash1l	Histone-lysine N-methyltransferase ash1
XLOC_027784	Epb41l3	Band 4.1-like protein 1
XLOC_027785	cora	coracle protein, putative
XLOC_027951	EaffTmpM007749	Excitatory amino acid transporter 1
XLOC_027989	Orct	Solute carrier family 22 member 6-A
XLOC_028060	ck	Myosin-VIIa
XLOC_028100	KIF21B	Kinesin-like protein KIF21B
XLOC_028101	Kif21a	Kinesin-like protein KIF21A
XLOC_028117	NCAN	Neurocan core protein
XLOC_028262	Clcn7	Chloride channel protein A
XLOC_028276	sll0108	Putative ammonium transporter sll0108
XLOC_028561	FGFRL1	Fibroblast growth factor receptor-like 1
XLOC_029421	EaffTmpM026627	Trypsin
XLOC_029424	EaffTmpM026628	PREDICTED: nocturnin isoform X1
XLOC_029654	loco	Regulator of G-protein signaling loco

XLOC_029806	Arsb	Arylsulfatase B
XLOC_029854	Adamts7	A disintegrin and metalloproteinase with thrombospondin motifs 7
XLOC_029945	BQ2027 MB0916	Uncharacterized monooxygenase Mb0916
XLOC_030077	EaffTmpS009797	RNA-directed DNA polymerase from mobile element jockey, partial
XLOC_030119	Ago1	Vesicle transport protein SEC20
XLOC_030142	lolal	Longitudinals lacking protein-like
XLOC_030143	lolal	Longitudinals lacking protein-like
XLOC_030321	alpha Man Ia	Mannosyl-oligosaccharide alpha-1,2-mannosidase isoform A
XLOC_030436	slmo	Protein slowmo
XLOC_030440	Nek8	Serine/threonine-protein kinase Nek8
XLOC_030441	nek3	Serine/threonine-protein kinase Nek8
XLOC_030443	EaffTmpM002282	predicted protein
XLOC_030452	PLA2G4A	Cytosolic phospholipase A2
XLOC_030461	Invadolysin	Leishmanolysin-like peptidase
XLOC_030562	Fas1	Fasciclin-1
XLOC_030564	FAS1	Fasciclin-1
XLOC_030565	FAS1	Fasciclin-1
XLOC_030566	Slc25a46	Solute carrier family 25 member 46
XLOC_030573	CASKIN1	Caskin-1
XLOC_030588	slc25a40	Solute carrier family 25 member 40
XLOC_030624	rap1b	Ras-related protein Rap-1b
XLOC_030948	EaffTmpM026454	transmembrane protein, putative
XLOC_030996	EaffTmpM026736	calmin-like protein
XLOC_031066	RPII	DNA-directed RNA polymerase II subunit RPB1
XLOC_031331	Cbx1	Chromobox protein homolog 1
XLOC_031451	ERMP1	Cysteine and glycine-rich protein 2
XLOC_031737	Myo1a	Unconventional myosin-Ia
XLOC_031738	MYO1A	Unconventional myosin-Ia
XLOC_031745	EaffTmpM028439	GG10482
XLOC_031829	EaffTmpM028878	CUG-BP- and ETR-3-like factor 1
XLOC_031921	EaffTmpM009431	Odr-4-like protein
XLOC_032116	IGDCC3	Immunoglobulin superfamily DCC subclass member 3
XLOC_032261	sll1483	Uncharacterized protein sll1483

XLOC_032364	Dmel Tmp cg9896	GH20916
XLOC_032605	EaffTmpM008215	putative SPT transcription factor family member
XLOC_032620	PCSK2	Neuroendocrine convertase 2
XLOC_032621	NHA-7	Na+/H+ Antiporter, paralog 7
XLOC_032632	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_032642	TkR86C	Tachykinin-like peptides receptor 86C
XLOC_032658	DUT	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial
XLOC_032661	Arsb	Arylsulfatase B
XLOC_032682	NHA-6	Na+/H+ Antiporter, paralog 6
XLOC_032683	NHA-5	Na+/H+ Antiporter, paralog 5
XLOC_032685	NHA-3	Na+/H+ Antiporter, paralog 3
XLOC_032687	NHA-1-frag	Na+/H+ Antiporter, paralog 1 (fragment)
XLOC_032688	NHA-1	Na+/H+ Antiporter, paralog 1
XLOC_032865	cac	Voltage-dependent calcium channel type A subunit alpha-1
XLOC_033548	NDUFS3	NADH dehydrogenase
XLOC_033549	NDUFS3	NADH dehydrogenase
XLOC_033551	DDX3X	ATP-dependent RNA helicase DDX3X
XLOC_033828	Smlt0970	Peptidyl-Asp metalloendopeptidase
XLOC_033832	osa	Trithorax group protein osa
XLOC_033833	osa	Trithorax group protein osa
XLOC_033911	EaffTmpS028717	xylose isomerase
XLOC_034131	ATP11A	Probable phospholipid-transporting ATPase IH
XLOC_034145	inx1	Innexin inx2
XLOC_034164	Prkg1	cGMP-dependent protein kinase 1
XLOC_034177	EaffTmpS009209	predicted protein
XLOC_034464	Rpl23	Nucleolar protein 6
XLOC_034640	SLC18B1	MFS-type transporter SLC18B1
XLOC_034641	EaffTmpM010522	chromaffin granule amine transporter, putative
XLOC_034707	EaffTmpM010552	AGAP004872-PA-like protein
XLOC_035036	trc	Serine/threonine-protein kinase tricorner
XLOC_035073	Hn	Protein henna
XLOC_035121	SCYL2	SCY1-like protein 2
XLOC_035122	Indy	Protein I'm not dead yet

XLOC_035193	DEPDC5	DEP domain-containing protein 5
XLOC_035467	KCNJ18	Inward rectifier potassium channel 18
XLOC_035541	NR5A2	Nuclear receptor subfamily 5 group A member 2
XLOC_035695	Tmprss11a	Transmembrane protease serine 11A
XLOC_035806	entpd7	Ectonucleoside triphosphate diphosphohydrolase 7
XLOC_035875	PDCD11	Protein RRP5 homolog
XLOC_035876	Pdcd11	Protein RRP5 homolog
XLOC_035882	FN1	PREDICTED: receptor-type tyrosine-protein phosphatase H
XLOC_035967	TNPO1	Transportin-1
XLOC_035968	Tnpo1	Transportin-1
XLOC_035975	MMD2	Monocyte to macrophage differentiation factor
XLOC_036011	qvr	Protein quiver
XLOC_036012	EaffTmpM013742	Pupal cuticle protein G1A, putative
XLOC_036013	bgm	Very long-chain-fatty-acid-CoA ligase bubblegum
XLOC_036014	acsbg2	Long-chain-fatty-acid-CoA ligase ACSBG1
XLOC_036016	EaffTmpM013752	Compound eye opsin BCRH2
XLOC_036026	Os10g0493600	Alpha-galactosidase
XLOC_036131	Cd63	CD63 antigen
XLOC_036449	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_036545	TY3B I	gag pol protein
XLOC_036576	EaffTmpM001762	hypothetical protein M91 04358, partial
XLOC_036940	EaffTmpM009997	GD20657
XLOC_037003	AASS	Alpha-aminoadipic semialdehyde synthase, mitochondrial
XLOC_037183	TTLL4	Tubulin polyglutamylase TTLL4
XLOC_037445	RPUSD2	RNA pseudouridylate synthase domain-containing protein 2
XLOC_037489	CA13	Carbonic anhydrase 13
XLOC_037519	CLNS1A	Putative all-trans-retinol 13,14-reductase
XLOC_037527	GATA4	GATA-binding factor C
XLOC_037572	CPNE8	Copine-8
XLOC_037796	DHAR4	Putative glutathione S-transferase DHAR4

Table S5. Candidate genes under selection detected in Lake Michigan with $PBE_{\_Window}$ against the ancestral saline populations (Montmagny and L'Isle Verte).

Gene ID	Gene Symbol	Description
XLOC_001703	to	Protein takeout
XLOC_001905	Taf7	Heat shock 70 kDa protein 4L
XLOC_002203	EaffTmpM013148	putative membrane protein
XLOC_002539	cxxc4	CXXC-type zinc finger protein 4
XLOC_002723	MM 2675	mCG117393, isoform CRA b
XLOC_002754	EaffTmpM011985	PREDICTED: type-2 ice-structuring protein-like, partial
XLOC_002834	Neurl4	Neuralized-like protein 4
XLOC_002893	OSBPL8	Oxysterol-binding protein-related protein 8
XLOC_002912	Inpp4a	Type I inositol-3,4-bisphosphate 4-phosphatase
XLOC_003022	PAPLN	Papilin
XLOC_003176	TMEM189	Transmembrane protein 189
XLOC_003274	ECE1	Neprilysin-1
XLOC_003455	Slc9a9	Sodium/hydrogen exchanger 9
XLOC_003457	nt5c2	Cytosolic purine 5'-nucleotidase
XLOC_003546	SLC27A1	Long-chain fatty acid transport protein 1
XLOC_003547	SLC27A4	Long-chain fatty acid transport protein 1
XLOC_003947	Fhdc1	FH2 domain-containing protein 1
XLOC_003948	EaffTmpM003234	formin 3
XLOC_004079	Nrt	Neurotactin
XLOC_004192	Pdzd2	PDZ domain-containing protein 2
XLOC_004445	Mhc	Myosin heavy chain, muscle
XLOC_004447	Gpr158	Probable G-protein coupled receptor 158
XLOC_004456	Dmel Tmp cg9395	UPF0392 protein F13G3.3
XLOC_004525	AAEL010189	Band 7 protein AAEL010189
XLOC_004652	Unc5b	Netrin receptor UNC5B
XLOC_004989	EaffTmpA002389	Probable transposable element

XLOC_004990	aldh8a1	Aldehyde dehydrogenase family 8 member A1
XLOC_004991	aldh8a1	Aldehyde dehydrogenase family 8 member A1
XLOC_005107	EaffTmpM002418	putative integral membrane protein
XLOC_005113	MEST	Mesoderm-specific transcript homolog protein
XLOC_005119	Iqgap2	Ras GTPase-activating-like protein IQGAP1
XLOC_005574	EaffTmpM015577	Macrophage MHC class I receptor 2-like protein
XLOC_005606	EaffTmpM015580	unnamed protein product
XLOC_005764	Lama1	Laminin subunit alpha-1
XLOC_005765	HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein
XLOC_005766	LAMA1	Laminin subunit alpha-1
XLOC_005767	Lama1	Laminin subunit alpha-2
XLOC_005771	EaffTmpM014173	PREDICTED: laminin subunit alpha-1-like
XLOC_005772	LAMA2	Laminin subunit alpha-2
XLOC_005775	AK	Arginine kinase
XLOC_005776	EaffTmpM014180	Protein lava lamp
XLOC_005802	EaffTmpM014144	collagen-binding protein A
XLOC_005811	EaffTmpS014176	ACYPI006308
XLOC_006148	ZNF142	Zinc finger protein 142
XLOC_006149	EaffTmpM016112	Venom carboxylesterase-6
XLOC_006180	Znf431	hypothetical protein LOTGIDRAFT 168530
XLOC_006342	EaffTmpS015194	Solute carrier family 35 member G1
XLOC_006370	HMCN1	Hemicentin-1
XLOC_006381	KLHL5	Kelch-like protein diablo
XLOC_006409	HMCN1	Hemicentin-1
XLOC_006427	hceb	High choriolytic enzyme 2
XLOC_006578	EaffTmpM016437	northern shrimp nuclease
XLOC_006767	Ddi2	Protein DDI1 homolog 2
XLOC_007207	EaffTmpM017845	Tubulin alpha-1 chain
XLOC_007262	Hsap Tmp gatad1	GATA zinc finger domain-containing protein 1, partial
XLOC_007263	gatad1	GATA zinc finger domain-containing protein 1
XLOC_007558	gpt2	Alanine aminotransferase 2
XLOC_007749	Best3	Bestrophin-3
XLOC_007829	Vps37b	Vacuolar protein sorting-associated protein 37B

XLOC_008012	ceh 9	Homeobox protein ceh-9
XLOC_008052	phyhd1	Phytanoyl-CoA dioxygenase domain-containing protein 1
XLOC_008127	tenm3	Teneurin-3
XLOC_008434	GluClalpha	Glutamate-gated chloride channel
XLOC_008492	NKA-a-5	Na+/K+-ATPase, subunit alpha, paralog 5
XLOC_008516	Akt1	RAC-gamma serine/threonine-protein kinase
XLOC_008619	ESRRG	Steroid hormone receptor ERR1
XLOC_008641	EaffTmpM004767	PREDICTED: uncharacterized protein LOC100893156
XLOC_008722	TCF25	Transcription factor 25
XLOC_009056	EaffTmpM018282	UPF0415 protein C7orf25 homolog
XLOC_009105	PUS3	Protein RCC2 homolog
XLOC_009350	NKA-b-4	Na+/K+-ATPase, subunit beta, paralog 4
XLOC_009363	CG12375	Beta-lactamase-like protein 2 homolog
XLOC_009447	Dmel Tmp akirin	Akirin-1
XLOC_009473	EaffTmpM003656	hypothetical protein DAPPUDRAFT 97956
XLOC_009810	TR2	Tropinone reductase 2
XLOC_010514	Plg	Serine proteinase stubble
XLOC_010515	RDH13	Retinol dehydrogenase 13
XLOC_010580	PSR	Bifunctional arginine demethylase and lysyl-hydroxylase PSR
XLOC_010639	ZNF628	Zinc finger protein 628
XLOC_010654	Dmel Tmp serp	AGAP011936-PA
XLOC_010667	EaffTmpM004495	glycosyltransferase C17G8.11c
XLOC_010670	TSPAN6	Tetraspanin-6
XLOC_011039	nAChRbeta2	Neuronal acetylcholine receptor subunit beta-4
XLOC_011040	nAChRbeta2	Acetylcholine receptor subunit beta-like 2
XLOC_011206	Elovl7	Elongation of very long chain fatty acids protein AAEL008004
XLOC_011451	pald1	Paladin
XLOC_011853	Klhl20	Sarcocystatin-A
XLOC_011954	eng1	Endo-1,3(4)-beta-glucanase 1
XLOC_011966	EaffTmpM017616	sodium/hydrogen exchanger-like domain-containing protein 1
XLOC_012007	EaffTmpM017618	PREDICTED: ras-related protein Rab-1B isoform X2
XLOC_012245	Mcm6	DNA replication licensing factor MCM6
XLOC_012593	EaffTmpM016928	Glutathione S-transferase

XLOC_012627	EaffTmpA018418	squash family serine protease inhibitor
XLOC_012643	CUL4B	Prostaglandin reductase 1
XLOC_012948	OVCH1	Ovochymase-1
XLOC_013215	Slc16a13	Monocarboxylate transporter 9
XLOC_013306	GstD1	Glutathione S-transferase 1, isoform D
XLOC_013461	spz	Protein spaetzle
XLOC_013473	Golga4	Golgin subfamily A member 4
XLOC_013757	Ac76E	Adenylate cyclase type 2
XLOC_014213	Dmel Tmp cg16787	Uncharacterized protein C6orf136 homolog
XLOC_014346	Siae	Sialate O-acetylesterase
XLOC_014393	Acsl3	Long-chain-fatty-acid-CoA ligase 3
XLOC_014422	PLCB4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-4
XLOC_014442	VDE1	Violaxanthin de-epoxidase, chloroplastic
XLOC_014449	flot1	Flotillin-1
XLOC_014455	CA-12	Carbonic Anhydrase, paralog 12
XLOC_014478	ncs 2	Neuronal calcium sensor 2
XLOC_014506	EaffTmpM022651	EF-hand domain-containing protein D1
XLOC_014637	NKA-a-2	Na+/K+-ATPase, subunit alpha, paralog 2
XLOC_014689	XDH	Xanthine dehydrogenase/oxidase
XLOC_014691	xdh	Xanthine dehydrogenase
XLOC_014692	Smlt0970	Peptidyl-Asp metalloendopeptidase
XLOC_014712	EaffTmpM020045	COG1292: Choline-glycine betaine transporter (ISS)
XLOC_014714	Smlt0970	Peptidyl-Asp metalloendopeptidase
XLOC_014758	C1qtnf4	Complement C1q tumor necrosis factor-related protein 4
XLOC_014781	Brf1	Transcription factor IIIB 90 kDa subunit
XLOC_014799	TkR99D	Tachykinin-like peptides receptor 99D
XLOC_014840	RpL4	60S ribosomal protein L4
XLOC_014850	EaffTmpM003935	Cysteine-rich motor neuron 1 protein
XLOC_014885	CG7708	High-affinity choline transporter 1
XLOC_014942	Slc26a11	Sodium-independent sulfate anion transporter
XLOC_014943	SLC26A11	Sodium-independent sulfate anion transporter
XLOC_015013	Nach	Na+ Channel
XLOC_015294	GRM6	Metabotropic glutamate receptor 8

XLOC_015390	Lman1	Protein ERGIC-53
XLOC_015393	Slit1	Slit homolog 1 protein
XLOC_015572	TALDO1	Transaldolase NQM1
XLOC_015599	UBE2J2	Ubiquitin-conjugating enzyme E2 J2
XLOC_015897	Mkx	Homeobox protein Mohawk
XLOC_015938	EaffTmpM021591	Protein roadkill
XLOC_016061	SULT1C4	Sulfotransferase 1C4
XLOC_016088	SHROOM2	AGAP008245-PA-like protein
XLOC_016209	Ephx1	Epoxide hydrolase 1
XLOC_016210	Ephx1	Epoxide hydrolase 1
XLOC_016230	TNT	Troponin T
XLOC_016250	Tmem131	Transmembrane protein 131
XLOC_016843	cut 1	Cuticlin-1
XLOC_016863	EaffTmpM022643	Saposin-like protein 11
XLOC_017082	EaffTmpM019891	63 kDa sperm flagellar membrane protein
XLOC_017183	Dmel Tmpssf8	GD21229
XLOC_017191	Tmprss9	Enteropeptidase
XLOC_017218	EaffTmpS022285	PREDICTED: uncharacterized protein LOC101237577
XLOC_017244	SVEP1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_017247	Svep1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_017248	CSMD2	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1, partial
XLOC_017971	Fuca	Putative alpha-L-fucosidase
XLOC_018191	rbck1	RanBP-type and C3HC4-type zinc finger-containing protein 1
XLOC_018302	PLB1	hypothetical protein PHAVU 007G184300g
XLOC_018307	CHIA	Acidic mammalian chitinase
XLOC_018517	EaffTmpS021888	Sarcoplasmic calcium-binding protein, beta chain
XLOC_018538	rhbg	Ammonium transporter Rh type B
XLOC_018539	KCNJ4	Inward rectifier potassium channel 4
XLOC_018540	KCNJ2	Inward rectifier potassium channel 2
XLOC_018666	EaffTmpM006183	hypothetical protein GUITHDRAFT 82324
XLOC_018725	Arid4b	at-rich interactive domain-containing protein 5B, putative
XLOC_018726	Arid4b	AT-rich interactive domain-containing protein 4B

XLOC_018727	Arid4b	AT-rich interactive domain-containing protein 4B
XLOC_018745	EaffTmpM006200	hypothetical protein LOTGIDRAFT 89004, partial
XLOC_018791	SRSF2	Serine/arginine-rich splicing factor 2
XLOC_018820	EaffTmpM006294	PREDICTED: zinc finger protein 778-like
XLOC_018833	ivns1abpb	Influenza virus NS1A-binding protein homolog B
XLOC_018857	EaffTmpM022124	Putative urea active transporter 1
XLOC_019421	Kif13a	Kinesin-like protein KIF13A
XLOC_019460	EaffTmpA022134	kelch-like protein 5
XLOC_019468	ROCK2	DNA-directed RNA polymerase II subunit RPB7
XLOC_019489	EaffTmpM024443	PREDICTED: uncharacterized protein LOC101743373
XLOC_019757	EaffTmpM007582	Alpha-parvin
XLOC_019788	AMT-6	Ammonia Transporter, paralog 6
XLOC_019811	ZFP64	RE1-silencing transcription factor A
XLOC_019830	Rpl3	60S ribosomal protein L3
XLOC_020005	Plscr1	Phospholipid scramblase 1
XLOC_020244	EaffTmpM007242	hypothetical protein THAOC 18358
XLOC_020267	Nme2	Nucleoside diphosphate kinase B
XLOC_020315	PKD2L1	Polycystic kidney disease 2-like 1 protein
XLOC_020410	MYO18A	Unconventional myosin-XVIIIa
XLOC_020532	DIRAS2	GTP-binding protein Di-Ras2
XLOC_020554	GABRP	Gamma-aminobutyric acid receptor subunit pi
XLOC_020604	mlt 7	Peroxidase mlt-7
XLOC_020818	pxdn	Peroxidasin
XLOC_020823	EaffTmpA006955	PREDICTED: uncharacterized protein LOC100883356
XLOC_020833	EaffTmpM006969	AGAP005830-PA-like protein
XLOC_020859	Hsap Tmp card6	putative integrase core domain protein
XLOC_020867	NKA-a-1	Na+/K+-ATPase, subunit alpha, paralog 1
XLOC_020897	5 HT1B	5-hydroxytryptamine receptor 2B
XLOC_021365	lact 2	Beta-lactamase domain-containing protein 2
XLOC_021369	EaffTmpM022895	Dynein heavy chain
XLOC_021435	EaffTmpM021841	Collagen alpha-2(IV) chain
XLOC_021452	Cpn1	Carboxypeptidase Z
XLOC_021463	ADAMTS6	A disintegrin and metalloproteinase with thrombospondin motifs 6

XLOC_021563	SMYD4	SET and MYND domain-containing protein 4
XLOC_021602	SPATA5L1	Spermatogenesis-associated protein 5-like protein 1
XLOC_021614	SRF	Serum response factor
XLOC_021803	EaffTmpM007399	conserved domain protein
XLOC_021828	Arsj	Arylsulfatase J
XLOC_021829	ARSJ	Arylsulfatase J
XLOC_022586	MANBA	Beta-mannosidase
XLOC_023176	EaffTmpM024767	hypothetical protein DAPPUDRAFT 311122
XLOC_023350	Srprb	Signal recognition particle receptor subunit beta
XLOC_023414	zip	Myosin heavy chain, non-muscle
XLOC_023579	Eip74EF	Ecdysone-induced protein 74EF isoform B
XLOC_023585	EaffTmpM024801	PREDICTED: ankyrin repeat domain-containing protein SOWAHB
XLOC_023592	EaffTmpS024819	Uncharacterized protein L116
XLOC_023790	PUM1	Pumilio homolog 1
XLOC_023882	Mrps22	28S ribosomal protein S22, mitochondrial
XLOC_023883	MRPS22	28S ribosomal protein S22, mitochondrial
XLOC_024192	EaffTmpM027123	Carbohydrate sulfotransferase 5, partial
XLOC_024209	Fubp1	Far upstream element-binding protein 1
XLOC_024279	CrebA	Cyclic AMP response element-binding protein A
XLOC_024283	EaffTmpS006313	PREDICTED: DNA-directed RNA polymerase II subunit RPB1-like isoform X2
XLOC_024284	TNT	Troponin T
XLOC_024337	EaffTmpA006399	Cysteine-rich, acidic integral membrane protein precursor, putative
XLOC_024502	EaffTmpM024342	hypothetical protein PPL 08161
XLOC_024517	GH17388	E3 UFM1-protein ligase 1 homolog
XLOC_024866	abr	Active breakpoint cluster region-related protein
XLOC_024895	Alk	Leukocyte tyrosine kinase receptor
XLOC_024909	Scyl2	SCY1-like protein 2
XLOC_024987	Sardh	Sarcosine dehydrogenase, mitochondrial
XLOC_025050	DLGAP4	Disks large-associated protein 4
XLOC_025220	EaffTmpA025565	Chymotrypsinogen A
XLOC_025339	CG18661	Longitudinals lacking protein-like
XLOC_025738	SLC24A5	Sodium/potassium/calcium exchanger 5
XLOC_025977	Paip1	Polyadenylate-binding protein-interacting protein 1

XLOC_026481	Bcs1l	Mitochondrial chaperone BCS1
XLOC_026484	kanE	Probable L-threonine 3-dehydrogenase
XLOC_026509	EaffTmpM008142	Venom carboxylesterase-6
XLOC_026536	ENAH	Protein enabled homolog
XLOC_026555	Glb1l2	Beta-galactosidase 17
XLOC_026615	Dmel Tmp cg13760	Protein GUCD1
XLOC_026688	Gid4	Sialin
XLOC_026998	ORF99	Putative apoptosis inhibitor ORF99
XLOC_027071	MYC	Myc protein
XLOC_027265	EaffTmpM008452	predicted protein
XLOC_027441	Ca9	Carbonic anhydrase 4
XLOC_027455	Wdr91	WD repeat-containing protein 91
XLOC_027457	CAV2	Caveolin-2
XLOC_027542	Ptpn12	Tyrosine-protein phosphatase non-receptor type 12
XLOC_027548	Bmp7	Bone morphogenetic protein 7
XLOC_027549	ZNF561	Zinc finger protein 561
XLOC_027662	EaffTmpM007993	Cuticle protein 8
XLOC_027781	SMYD4	SET and MYND domain-containing protein 4
XLOC_027930	FHL2	Four and a half LIM domains protein 2
XLOC_027951	EaffTmpM007749	Excitatory amino acid transporter 1
XLOC_027955	Mmp9	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_027978	Taf9b	Signal recognition particle 9 kDa protein
XLOC_028276	sll0108	Putative ammonium transporter sll0108
XLOC_028537	MAP6	ATP synthase lipid-binding protein, mitochondrial
XLOC_028709	ncs1	Calcium-binding protein NCS-1
XLOC_028731	tmem129	E3 ubiquitin-protein ligase TM129
XLOC_029135	PIM3	Serine/threonine-protein kinase pim-3
XLOC_029253	EaffTmpS009329	predicted protein
XLOC_029254	EaffTmpM009330	SPT transcription factor family member
XLOC_029331	FucTC	hypothetical protein DAPPUDRAFT 241186
XLOC_029387	Tre1	Protein trapped in endoderm-1
XLOC_029594	pbo 4	Na(+)/H(+) exchanger protein 7

XLOC_029640	Abcc10	Protein VAC14 homolog
XLOC_029786	Sqstm1	Sequestosome-1
XLOC_030077	EaffTmpS009797	RNA-directed DNA polymerase from mobile element jockey, partial
XLOC_030078	Htatsf1	HIV Tat-specific factor 1 homolog
XLOC_030079	Slc35c2	Solute carrier family 35 member C2
XLOC_030142	lolal	Longitudinals lacking protein-like
XLOC_030143	lolal	Longitudinals lacking protein-like
XLOC_030190	EaffTmpM027647	PR domain zinc finger protein 10
XLOC_030199	EaffTmpM027648	merozoite surface protein 9, partial
XLOC_030461	Invadolysin	Leishmanolysin-like peptidase
XLOC_030540	Gxylt2	Glucoside xylosyltransferase 2
XLOC_030570	tdh	L-threonine 3-dehydrogenase
XLOC_030997	NKCC2	Na+,K+,2Cl- Cotransporter, paralog 2
XLOC_031011	TCNA	Sialidase
XLOC_031012	MYLK	Myosin light chain kinase, smooth muscle
XLOC_031013	EaffTmpM028461	Snake venom metalloprotease inhibitor 02A10, partial
XLOC_031352	SEPSECS	O-phosphoseryl-tRNA(Sec) selenium transferase
XLOC_031862	mesh	Protein mesh
XLOC_031863	mesh	Protein mesh
XLOC_031865	EaffTmpM009345	Protein mesh
XLOC_031867	EaffTmpM009347	Protein mesh
XLOC_032137	pou2f1	POU domain, class 2, transcription factor 1
XLOC_032142	slo	Calcium-activated potassium channel slowpoke
XLOC_032144	slo	Calcium-activated potassium channel slowpoke
XLOC_032347	OIT3	Uromodulin
XLOC_032490	Orct	Organic cation transporter protein
XLOC_032497	Hex t2	Hexokinase type 2
XLOC_032544	Ufl1	E3 UFM1-protein ligase 1
XLOC_032598	Inpp5e	72 kDa inositol polyphosphate 5-phosphatase
XLOC_032620	PCSK2	Neuroendocrine convertase 2
XLOC_032632	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_032668	Tret1 2	Facilitated trehalose transporter Tret1
XLOC_032682	NHA-6	Na+/H+ Antiporter, paralog 6

XLOC_032683	NHA-5	Na+/H+ Antiporter, paralog 5
XLOC_032684	NHA-4	Na+/H+ Antiporter, paralog 4
XLOC_032685	NHA-3	Na+/H+ Antiporter, paralog 3
XLOC_032687	NHA-1-frag	Na+/H+ Antiporter, paralog 1 (fragment)
XLOC_032688	NHA-1	Na+/H+ Antiporter, paralog 1
XLOC_032863	ninaB	Carotenoid isomerooxygenase
XLOC_032864	EaffTmpM010258	Carotenoid cleavage dioxygenase 8 homolog A, chloroplastic
XLOC_033428	mthl6	Probable G-protein coupled receptor Mth-like 6
XLOC_033990	Pcnx2	Pecanex-like protein 1
XLOC_034032	EaffTmpM011015	hypothetical protein BRAFLDRAFT 220632
XLOC_034313	PIM3	Serine/threonine-protein kinase pim-3
XLOC_034337	Dmel Tmp ewg	DNA-binding protein P3A2
XLOC_034347	SMYD4	SET and MYND domain-containing protein DDB G0292140
XLOC_034641	EaffTmpM010522	chromaffin granule amine transporter, putative
XLOC_034663	EaffTmpM010554	hypothetical protein L798 10568
XLOC_034707	EaffTmpM010552	AGAP004872-PA-like protein
XLOC_034709	EaffTmpM010553	PREDICTED: uncharacterized protein LOC100118488
XLOC_034902	AP3B2	Ran-specific GTPase-activating protein
XLOC_034958	EaffTmpM001953	Thioredoxin H1
XLOC_034959	Slc6a7	Sodium-dependent proline transporter
XLOC_035124	EaffTmpM002051	hypothetical protein H312 02030
XLOC_035138	NELL1	Protein kinase C-binding protein NELL1
XLOC_035310	EaffTmpM029131	LITAF homolog
XLOC_035422	EaffTmpS011169	AGAP004872-PA-like protein
XLOC_035435	COPS6	COP9 signalosome complex subunit 6
XLOC_035467	KCNJ18	Inward rectifier potassium channel 18
XLOC_035570	Cyp301a1	Probable cytochrome P450 301a1, mitochondrial
XLOC_035588	AP1G1	AP-1 complex subunit gamma-1
XLOC_035810	PGAP1	GPI inositol-deacylase
XLOC_035812	PGAP1	GPI inositol-deacylase
XLOC_035870	thbs3b	Thrombospondin-1
XLOC_035872	Cele Tmp w03g9.7	uncharacterized protein LOC576686 isoform 1
XLOC_035874	RpLP0	60S acidic ribosomal protein P2

XLOC_035875	PDCD11	Protein RRP5 homolog
XLOC_036139	DHCR24	Delta(24)-sterol reductase
XLOC_036140	DHCR24	Delta(24)-sterol reductase
XLOC_036174	pif1	ATP-dependent DNA helicase PIF1
XLOC_036371	EaffTmpM011192	DOMON domain containing protein
XLOC_036449	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_037059	SMYD4	SET and MYND domain-containing protein 4
XLOC_037330	EaffTmpS011905	unnamed protein product
XLOC_037359	abhd12	Uncharacterized protein slr1819
XLOC_037427	EaffTmpM029200	E3 ubiquitin-protein ligase sina
XLOC_037601	EaffTmpM012205	hypothetical protein BRAFLDRAFT 63701
XLOC_037773	CG10336	Protein TIPIN homolog
XLOC_037774	Sfmbt1	Scm-like with four MBT domains protein 1
XLOC_037777	Skeletor	Protein Skeletor, isoforms B/C
XLOC_037782	CEP290	Centrosomal protein of 290 kDa
XLOC_037799	EaffTmpM010914	Cuticle protein 7
XLOC_037810	Cubn	Cubilin
XLOC_037829	SLC12A6	Solute carrier family 12 member 5

Table S6. Candidate genes under selection detected in Lake Michigan with $PBE_{\_MaxSNP}$ against the ancestral saline populations (Montmagny and L'Isle Verte).

Gene ID	**Gene Symbol**	**Description**
XLOC_000314	Csk	Tyrosine-protein kinase CSK
XLOC_000602	Dmel Tmp itp	Ion transport peptide
XLOC_000629	EaffTmpA002714	Protein unzipped
XLOC_000709	Gbeta76C	Guanine nucleotide-binding protein subunit beta-2
XLOC_000874	Fkbp6	Inactive peptidyl-prolyl cis-trans isomerase FKBP6
XLOC_000905	Dmel Tmp mas	Plasma kallikrein

XLOC_001036	APOD	Apolipoprotein D
XLOC_001043	EaffTmpM012603	hypothetical protein
XLOC_001052	EaffTmpM012607	Serine-aspartate repeat-containing protein I
XLOC_001055	CA-14	Carbonic Anhydrase, paralog 14
XLOC_001068	SFXN1	Sideroflexin-1
XLOC_001069	HtrA2	Serine protease HTRA2, mitochondrial
XLOC_001070	mtp 18	Mitochondrial fission process protein 1
XLOC_001095	MYLK	Myosin light chain kinase, smooth muscle
XLOC_001101	THAP9	DNA transposase THAP9
XLOC_001105	EaffTmpM012611	RNA-binding protein 12B
XLOC_001115	VDE1	Violaxanthin de-epoxidase, chloroplastic
XLOC_001132	NAXD	ATP-dependent (S)-NAD(P)H-hydrate dehydratase
XLOC_001133	Roe1	GrpE protein homolog, mitochondrial
XLOC_002256	EaffTmpM012429	Cholinesterase 2
XLOC_002268	EaffTmpA012442	hypothetical protein
XLOC_002308	RPGR	hypothetical protein
XLOC_002310	GSN	Gelsolin
XLOC_002328	mec 2	Band 7 protein AGAP004871
XLOC_002357	Orct	Organic cation transporter protein
XLOC_002615	EaffTmpM013442	calmin-like protein
XLOC_002754	EaffTmpM011985	PREDICTED: type-2 ice-structuring protein-like, partial
XLOC_002801	EaffTmpM003472	Protein Bm3600, isoform d
XLOC_002821	srebf2	Sterol regulatory element-binding protein 2
XLOC_003007	EaffTmpM012972	Flagellar calcium-binding protein
XLOC_003049	slc36a4	Proton-coupled amino acid transporter 3, partial
XLOC_003050	PLA2G1B	Phospholipase A2
XLOC_003218	Capr	Caprin homolog
XLOC_003270	RpL28	60S ribosomal protein L28
XLOC_003343	ESYT3	Probable ADP-ribosylation factor GTPase-activating protein AGD11
XLOC_003423	Zbtb41	Zinc finger and BTB domain-containing protein 41
XLOC_003440	caup	Homeobox protein caupolican
XLOC_003455	Slc9a9	Sodium/hydrogen exchanger 9
XLOC_003799	SLC13A5	Solute carrier family 13 member 5

XLOC_003973	FUT1	Galactoside 2-alpha-L-fucosyltransferase 1
XLOC_003974	EaffTmpM003115	Protein K04A8.1
XLOC_004013	Dis3l	DIS3-like exonuclease 1
XLOC_004014	dis3l	DIS3-like exonuclease 1
XLOC_004036	fax	Failed axon connections
XLOC_004157	EaffTmpM014578	BRCA1-associated protein
XLOC_004158	CIR1	Corepressor interacting with RBPJ 1
XLOC_004162	F40A3.3	Phosphatidylethanolamine-binding protein homolog F40A3.3
XLOC_004163	OV16	Putative odorant-binding protein A5
XLOC_004344	sgsm3	Small G protein signaling modulator 3
XLOC_004345	sgsm3	Small G protein signaling modulator 3
XLOC_004455	EaffTmpM012062	conserved hypothetical protein
XLOC_004456	Dmel Tmp cg9395	UPF0392 protein F13G3.3
XLOC_004637	EaffTmpM015926	hypothetical protein LOTGIDRAFT 166205
XLOC_004822	sno1	Senecionine N-oxygenase
XLOC_004863	b9d2	B9 domain-containing protein 2
XLOC_004873	POLR3C	DNA-directed RNA polymerase III subunit RPC3
XLOC_004901	Bhmt	predicted protein
XLOC_004987	Acsl1	Long-chain-fatty-acid-CoA ligase 5
XLOC_005020	ImpL2	Neural/ectodermal development factor IMP-L2
XLOC_005107	EaffTmpM002418	putative integral membrane protein
XLOC_005112	Pgk	Phosphoglycerate kinase
XLOC_005113	MEST	Mesoderm-specific transcript homolog protein
XLOC_005194	Hnrnpul1	Heterogeneous nuclear ribonucleoprotein U-like protein 1
XLOC_005587	Pkd1l2	Polycystic kidney disease protein 1-like 2
XLOC_005710	nhr 41	Nuclear hormone receptor family member nhr-41
XLOC_005725	EaffTmpA014069	N-acylglucosamine 2-epimerase
XLOC_005771	EaffTmpM014173	PREDICTED: laminin subunit alpha-1-like
XLOC_005772	LAMA2	Laminin subunit alpha-2
XLOC_005775	AK	Arginine kinase
XLOC_005776	EaffTmpM014180	Protein lava lamp
XLOC_005785	ASB2	Ankyrin repeat and SOCS box protein 2
XLOC_005811	EaffTmpS014176	ACYPI006308

XLOC_005817	haao	3-hydroxyanthranilate 3,4-dioxygenase
XLOC_005839	FAM46A	Protein FAM46A
XLOC_006149	EaffTmpM016112	Venom carboxylesterase-6
XLOC_006169	QSOX1	Sulfhydryl oxidase 1
XLOC_006252	CD109	CD109 antigen
XLOC_006387	EaffTmpA013269	calmin-like protein
XLOC_006408	HMCN1	Hemicentin-1
XLOC_006409	HMCN1	Hemicentin-1
XLOC_006412	RUFY2	tRNA (guanine(26)-N(2))-dimethyltransferase
XLOC_006427	hceb	High choriolytic enzyme 2
XLOC_006578	EaffTmpM016437	northern shrimp nuclease
XLOC_006897	AH9.1	Probable G-protein coupled receptor AH9.1
XLOC_006994	SPBPJ4664.02	predicted protein
XLOC_007003	Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit
XLOC_007015	SRP72	Signal recognition particle subunit SRP72
XLOC_007021	EaffTmpM016693	hypothetical protein Phum PHUM334420
XLOC_007030	GCDH	Glutaryl-CoA dehydrogenase, mitochondrial
XLOC_007207	EaffTmpM017845	Tubulin alpha-1 chain
XLOC_007261	AAEL003512	Aquaporin AQPAe.a
XLOC_007558	gpt2	Alanine aminotransferase 2
XLOC_007740	PTPRB	Receptor-type tyrosine-protein phosphatase F
XLOC_007743	ceh 9	Homeobox protein ceh-9
XLOC_007749	Best3	Bestrophin-3
XLOC_007780	Nach-PPK28	Pickpocket protein 28
XLOC_007831	Mcoln3	Mucolipin-3
XLOC_007987	EaffTmpM017116	AGAP012241-PA
XLOC_007988	EaffTmpM017117	GH16343
XLOC_008050	Ttpal	Retinaldehyde-binding protein 1-like protein 1
XLOC_008063	Fbxo21	F-box only protein 21
XLOC_008090	KIF18B	Kinesin-like protein KIF18A
XLOC_008120	Cd36	Platelet glycoprotein 4
XLOC_008129	Dll1	Delta-like protein 1
XLOC_008434	GluClalpha	Glutamate-gated chloride channel

XLOC_008479	msta	Protein msta, isoform B
XLOC_008492	NKA-a-5	Na+/K+-ATPase, subunit alpha, paralog 5
XLOC_008581	EaffTmpM004812	Neuropilin-1, partial
XLOC_008619	ESRRG	Steroid hormone receptor ERR1
XLOC_008722	TCF25	Transcription factor 25
XLOC_008762	msta	Protein msta, isoform A
XLOC_008842	PIM3	Serine/threonine-protein kinase pim-3
XLOC_008891	Rpn6	26S proteasome non-ATPase regulatory subunit 11
XLOC_009137	GABRR2	Gamma-aminobutyric acid receptor subunit rho-2
XLOC_009286	Pxn	Peroxidasin
XLOC_009350	NKA-b-4	Na+/K+-ATPase, subunit beta, paralog 4
XLOC_009357	Pard6g	Partitioning defective 6 homolog gamma
XLOC_009358	EaffTmpM003636	GG15058
XLOC_009453	EaffTmpM003635	AGAP004872-PA
XLOC_009461	Xpo1	Exportin-1
XLOC_009568	EaffTmpS017377	GA14455
XLOC_009623	EaffTmpM016249	Hypothetical protein CBG24990
XLOC_009624	DDB G0290685	Uncharacterized protein DDB G0290685
XLOC_010574	Dmel Tmp verm	AGAP011937-PA-like protein
XLOC_010654	Dmel Tmp serp	AGAP011936-PA
XLOC_010667	EaffTmpM004495	glycosyltransferase C17G8.11c
XLOC_010670	TSPAN6	Tetraspanin-6
XLOC_011040	nAChRbeta2	Acetylcholine receptor subunit beta-like 2
XLOC_011178	EaffTmpM018062	cell surface protein
XLOC_011206	Elovl7	Elongation of very long chain fatty acids protein AAEL008004
XLOC_011250	EaffTmpM018315	hypothetical protein
XLOC_011301	RDH12	Retinol dehydrogenase 13
XLOC_011311	Prss41	Serine protease 41
XLOC_011377	CLS	Probable cardiolipin synthase (CMP-forming)
XLOC_011379	sucg 1	Probable succinyl-CoA ligase
XLOC_011415	EaffTmpM005193	serine protease
XLOC_011416	AQP3	Aquaporin-3
XLOC_011739	Slc2a1	Solute carrier family 2, facilitated glucose transporter member 1

XLOC_011751	Hex t2	Hexokinase type 2
XLOC_011752	EaffTmpA019597	hypothetical protein BRAFLDRAFT 117193
XLOC_011843	AGAP3	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3
XLOC_011853	Klhl20	Sarcocystatin-A
XLOC_011921	EaffTmpS018476	SET and MYND domain-containing protein 4
XLOC_012156	Map3k7	Mitogen-activated protein kinase kinase kinase 7
XLOC_012627	EaffTmpA018418	squash family serine protease inhibitor
XLOC_012733	EaffTmpM019915	Dipeptidyl aminopeptidase-like protein 6
XLOC_013137	chico	Semaphorin-1A
XLOC_013395	EaffTmpM019442	Pleckstrin domain-containing family O member 1, partial
XLOC_013633	Phf19	PHD finger protein 19
XLOC_013655	RP1L1	Retinitis pigmentosa 1-like 1 protein
XLOC_013656	EaffTmpM021021	Galactoside 2-alpha-L-fucosyltransferase 1
XLOC_013674	EaffTmpM019126	hypothetical protein
XLOC_013689	Nf1	Neurofibromin
XLOC_013718	ZDHHC14	Probable palmitoyltransferase ZDHHC14
XLOC_013726	prrc1	Protein PRRC1
XLOC_013729	EaffTmpA020123	AF308673 2 cell surface mucin-like protein
XLOC_013742	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013747	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013748	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013749	DNAJC13	DnaJ homolog subfamily C member 13
XLOC_013757	Ac76E	Adenylate cyclase type 2
XLOC_014213	Dmel Tmp cg16787	Uncharacterized protein C6orf136 homolog
XLOC_014393	Acsl3	Long-chain-fatty-acid-CoA ligase 3
XLOC_014400	bicc1 b	Protein bicaudal C homolog 1-B
XLOC_014434	Pabpn1l b	Embryonic polyadenylate-binding protein 2-B
XLOC_014442	VDE1	Violaxanthin de-epoxidase, chloroplastic
XLOC_014455	CA-12	Carbonic Anhydrase, paralog 12
XLOC_014603	LUC7L	Putative RNA-binding protein Luc7-like 1
XLOC_014604	RAB5B	Ras-related protein Rab-5C
XLOC_014612	EaffTmpM019533	Microtubule-associated protein tau

XLOC_014637	NKA-a-2	Na+/K+-ATPase, subunit alpha, paralog 2
XLOC_014689	XDH	Xanthine dehydrogenase/oxidase
XLOC_014713	betT	High-affinity choline transport protein
XLOC_014714	Smlt0970	Peptidyl-Asp metalloendopeptidase
XLOC_014829	TBCEL	Tubulin-specific chaperone cofactor E-like protein
XLOC_014840	RpL4	60S ribosomal protein L4
XLOC_014864	byn	T-related protein
XLOC_014880	nrf 6	Nose resistant to fluoxetine protein 6
XLOC_014915	EaffTmpM020780	Troponin C, isoform 2A
XLOC_014932	SKIV2L	Helicase SKI2W
XLOC_015007	mpp7	MAGUK p55 subfamily member 7
XLOC_015009	vps33b	Vacuolar protein sorting-associated protein 33B
XLOC_015013	Nach	Na+ Channel
XLOC_015021	Itpkb	Inositol-trisphosphate 3-kinase B
XLOC_015023	Sir2	Lysine-specific demethylase 2B
XLOC_015176	ENPEP	Glutamyl aminopeptidase
XLOC_015178	eIF3 S8	Eukaryotic translation initiation factor 3 subunit C
XLOC_015191	EaffTmpM021062	AGAP002367-PA
XLOC_015317	gata3	GATA-binding factor 3
XLOC_015598	EaffTmpM021462	C. briggsae CBR-ABT-1 protein
XLOC_015615	Gcnt3	Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 3
XLOC_015686	RYK	Tyrosine-protein kinase RYK
XLOC_015950	CCAP R	Cardioacceleratory peptide receptor
XLOC_015983	Sb	Serine proteinase stubble
XLOC_016061	SULT1C4	Sulfotransferase 1C4
XLOC_016135	TTC7B	tetratricopeptide repeat protein, tpr
XLOC_016137	EaffTmpM006131	Tetratricopeptide repeat protein 7B, partial
XLOC_016209	Ephx1	Epoxide hydrolase 1
XLOC_016210	Ephx1	Epoxide hydrolase 1
XLOC_016213	EaffTmpM022085	hypothetical protein M514 27541
XLOC_016219	Runx3	Runt-related transcription factor 3
XLOC_016220	Pitpnb	Phosphatidylinositol transfer protein beta isoform
XLOC_016228	Dapk1	Death-associated protein kinase 1

XLOC_016230	TNT	Troponin T
XLOC_016255	Chst5	Carbohydrate sulfotransferase 5
XLOC_016256	Zasp52	PDZ and LIM domain protein Zasp
XLOC_016275	EaffTmpM022237	hypothetical protein
XLOC_016276	Apaf1	Apoptotic protease-activating factor 1
XLOC_016281	EaffTmpM022244	Transmembrane protein 20
XLOC_016303	Slc4a4	Electrogenic sodium bicarbonate cotransporter 1
XLOC_016305	Slc4a4	Electrogenic sodium bicarbonate cotransporter 1
XLOC_016345	EaffTmpM021398	PREDICTED: uncharacterized protein LOC103569710
XLOC_016350	Nckx30C	Sodium/potassium/calcium exchanger Nckx30C
XLOC_016791	EaffTmpM022606	PREDICTED: uncharacterized protein CG7065-like isoform X1
XLOC_016804	EaffTmpM022606	hypothetical protein BRAFLDRAFT 195927
XLOC_016938	PLA2G7	Platelet-activating factor acetylhydrolase
XLOC_016939	Ttc8	Tetratricopeptide repeat protein 8
XLOC_017006	EaffTmpM020977	PREDICTED: uncharacterized protein LOC100892058
XLOC_017031	Harbi1	Putative nuclease HARBI1
XLOC_017032	GLRX	Glutaredoxin-1
XLOC_017284	cdk4	Cyclin-dependent kinase 4
XLOC_017307	DENND4A	C-myc promoter-binding protein
XLOC_017308	DENND4B	DENN domain-containing protein 4B
XLOC_017334	ZNF208	Zinc finger protein 681
XLOC_017358	Samd5	Sterile alpha motif domain-containing protein 5
XLOC_017371	Cpsf2	Cleavage and polyadenylation specificity factor subunit 2
XLOC_017447	Dclre1c	Protein artemis
XLOC_017450	marf1	Meiosis arrest female protein 1 homolog
XLOC_017453	EHF	ETS homologous factor
XLOC_017454	suz12b	Polycomb protein suz12-B
XLOC_017470	ADAM9	Glucose-dependent insulinotropic receptor
XLOC_017483	PDF	Transforming protein Qin
XLOC_017488	rmdn2	Regulator of microtubule dynamics protein 2
XLOC_017489	rmdn3	Regulator of microtubule dynamics protein 3
XLOC_017829	EaffTmpM022959	multiple banded antigen
XLOC_017845	C2orf16	Uncharacterized protein C2orf16

XLOC_017896	Kcnip1	Kv channel-interacting protein 1
XLOC_017971	Fuca	Putative alpha-L-fucosidase
XLOC_018302	PLB1	hypothetical protein PHAVU 007G184300g
XLOC_018409	MYLK	circumsporozoite protein
XLOC_018516	Mrc1	Macrophage mannose receptor 1, partial
XLOC_018517	EaffTmpS021888	Sarcoplasmic calcium-binding protein, beta chain
XLOC_018524	CD63	CD63 antigen
XLOC_018538	rhbg	Ammonium transporter Rh type B
XLOC_018539	KCNJ4	Inward rectifier potassium channel 4
XLOC_018540	KCNJ2	Inward rectifier potassium channel 2
XLOC_018575	EaffTmpM025025	GJ17509
XLOC_018696	Inx2	Innexin inx2
XLOC_018710	RRBP1	Mannan-binding lectin serine protease 2
XLOC_018791	SRSF2	Serine/arginine-rich splicing factor 2
XLOC_018811	EaffTmpM006286	predicted protein
XLOC_019383	ABCB6	ATP-binding cassette sub-family B member 6, mitochondrial
XLOC_019410	Gem	GTP-binding protein GEM
XLOC_019460	EaffTmpA022134	kelch-like protein 5
XLOC_019468	ROCK2	DNA-directed RNA polymerase II subunit RPB7
XLOC_019675	Inx2	Innexin inx2
XLOC_019830	Rpl3	60S ribosomal protein L3
XLOC_019928	pxt	Chorion peroxidase
XLOC_020036	SPNS2	Protein spinster homolog 2
XLOC_020072	Plcd4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-3-A
XLOC_020095	dcr 1	Endoribonuclease Dcr-1
XLOC_020096	DCL1	Endoribonuclease Dicer homolog 1
XLOC_020270	EaffTmpM007164	PREDICTED: hypothetical protein LOC100645147
XLOC_020345	EaffTmpM016330	Trypsin-1
XLOC_020410	MYO18A	Unconventional myosin-XVIIIa
XLOC_020532	DIRAS2	GTP-binding protein Di-Ras2
XLOC_020604	mlt 7	Peroxidase mlt-7
XLOC_020740	XCC0955	Peptidyl-Asp metalloendopeptidase
XLOC_020744	Slc16a5	Protein LIAT1

XLOC_020823	EaffTmpA006955	PREDICTED: uncharacterized protein LOC100883356
XLOC_020867	NKA-a-1	Na+/K+-ATPase, subunit alpha, paralog 1
XLOC_020894	SSR2	Translocon-associated protein subunit beta
XLOC_020932	Cyp2j6	Cytochrome P450 2J6
XLOC_021055	AHI1	Jouberin
XLOC_021061	Plscr2	Phospholipid scramblase 2
XLOC_021278	dnc	PREDICTED: cAMP-specific 3',5'-cyclic phosphodiesterase-like isoform 2
XLOC_021301	EaffTmpM005341	putative glutamine rich 2-like isoform 1
XLOC_021369	EaffTmpM022895	Dynein heavy chain
XLOC_021654	Nt5e	5'-nucleotidase
XLOC_021661	EaffTmpM022000	PREDICTED: uncharacterized protein LOC100167668, partial
XLOC_021757	EaffTmpM007444	hypothetical protein BCAMP 12085, partial
XLOC_021810	HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1
XLOC_021825	KLKB1	Plasma kallikrein
XLOC_022242	CG11007	Thioredoxin-related transmembrane protein 2 homolog
XLOC_022427	EaffTmpA026070	putative juvenile hormone esterase
XLOC_022586	MANBA	Beta-mannosidase
XLOC_022972	XYLB	Xylulose kinase
XLOC_023071	EaffTmpM024527	FAD-binding type 2
XLOC_023203	mdh	Malate dehydrogenase
XLOC_023350	Srprb	Signal recognition particle receptor subunit beta
XLOC_023352	slc38a7	Putative sodium-coupled neutral amino acid transporter 7
XLOC_023363	EaffTmpM025003	hypothetical protein DAPPUDRAFT 327174
XLOC_023400	Slc13a3	Solute carrier family 13 member 3
XLOC_023666	Slc18b1	MFS-type transporter SLC18B1
XLOC_023702	EaffTmpM026582	TMEM9 family protein
XLOC_023944	EaffTmpM025801	Tubulin alpha-1 chain
XLOC_023974	Grn	Granulins
XLOC_024209	Fubp1	Far upstream element-binding protein 1
XLOC_024216	CDC20	Cell division cycle protein 20 homolog
XLOC_024217	EaffTmpM006323	Tropomyosin
XLOC_024274	EaffTmpS006300	Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1
XLOC_024283	EaffTmpS006313	PREDICTED: DNA-directed RNA polymerase II subunit RPB1-like

		isoform X2
XLOC_024322	EaffTmpM006382	Sodium channel protein Nach
XLOC_024337	EaffTmpA006399	Cysteine-rich, acidic integral membrane protein precursor, putative
XLOC_024347	EaffTmpM006409	PDZ and LIM domain protein 3
XLOC_024367	Dnah3	Dynein heavy chain 3, axonemal
XLOC_024375	DNAH3	Dynein heavy chain 3, axonemal
XLOC_024528	EaffTmpS026560	hypothetical protein CGI 10027079
XLOC_024630	EaffTmpM025852	uncharacterized protein LOC733261
XLOC_024821	CtBP	C-terminal-binding protein
XLOC_024856	AGMO	Alkylglycerol monooxygenase
XLOC_024866	abr	Active breakpoint cluster region-related protein
XLOC_024867	FBXL20	F-box/LRR-repeat protein 20
XLOC_024868	CHRNA7	Neuronal acetylcholine receptor subunit alpha-7
XLOC_024987	Sardh	Sarcosine dehydrogenase, mitochondrial
XLOC_025009	Mafg	Transcription factor MafG
XLOC_025050	DLGAP4	Disks large-associated protein 4
XLOC_025073	Plxna4	Plexin-A4
XLOC_025074	PLXNA4	Plexin-A4
XLOC_025115	PEG3	Paternally-expressed gene 3 protein
XLOC_025116	EaffTmpM009100	predicted protein
XLOC_025124	VhaAC39 1	V-type proton ATPase subunit d 1
XLOC_025130	EaffTmpS009119	Huntingtin-interacting protein
XLOC_025132	HIP1	Huntingtin-interacting protein 1
XLOC_025168	EaffTmpM025111	Muscle calcium channel subunit alpha-1
XLOC_025204	Lcch3	Gamma-aminobutyric acid receptor subunit beta-like
XLOC_025220	EaffTmpA025565	Chymotrypsinogen A
XLOC_025379	NFE2L1	PHIST domain containing protein
XLOC_025589	Map3k4	Mitogen-activated protein kinase kinase kinase 4
XLOC_025891	Ces4a	Carboxylesterase 4A
XLOC_025907	MMP14	Octopamine receptor
XLOC_025933	EaffTmpM009049	Probable nitrile hydratase
XLOC_026124	Gyc88E	Soluble guanylate cyclase 88E
XLOC_026198	SLC4A11	Sodium bicarbonate transporter-like protein 11

XLOC_026420	ZW	Glucose-6-phosphate 1-dehydrogenase
XLOC_026462	REXO1L1P	Putative exonuclease GOR
XLOC_026468	EaffTmpM026671	Bov11.b1
XLOC_026481	Bcs1l	Mitochondrial chaperone BCS1
XLOC_026536	ENAH	Protein enabled homolog
XLOC_026615	Dmel Tmp cg13760	Protein GUCD1
XLOC_026688	Gid4	Sialin
XLOC_026711	EaffTmpS025892	DNA-dependent RNA polymerase II largest subunit, partial
XLOC_026873	EaffTmpM026295	AGAP004872-PA
XLOC_026958	Ptch2	Protein patched homolog 2
XLOC_026960	NPC1	Patched domain-containing protein 3
XLOC_027231	impact B	Telomerase reverse transcriptase
XLOC_027371	X element\ORF2	Probable RNA-directed DNA polymerase from transposon X-element
XLOC_027455	Wdr91	WD repeat-containing protein 91
XLOC_027456	EaffTmpM027089	Kelch motif, partial
XLOC_027457	CAV2	Caveolin-2
XLOC_027533	EaffTmpS027137	Myomodulin neuropeptides 1
XLOC_027549	ZNF561	Zinc finger protein 561
XLOC_027662	EaffTmpM007993	Cuticle protein 8
XLOC_027690	shg	DE-cadherin
XLOC_027792	reep4	Receptor expression-enhancing protein 4
XLOC_027808	Tmed3	Transmembrane emp24 domain-containing protein 3
XLOC_027951	EaffTmpM007749	Excitatory amino acid transporter 1
XLOC_027955	Mmp9	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_028098	IVL	Involucrin
XLOC_028110	EaffTmpM026815	Zinc finger protein 346
XLOC_028117	NCAN	Neurocan core protein
XLOC_028330	ankrd39	Ankyrin repeat domain-containing protein 39 homolog
XLOC_028437	Slc36a2	Proton-coupled amino acid transporter 2
XLOC_028446	EaffTmpM027454	cuticular protein 27 precursor
XLOC_028480	IFI30	Gamma-interferon-inducible lysosomal thiol reductase
XLOC_028491	GLRA2	Glycine receptor subunit alpha-2

XLOC_028709	ncs1	Calcium-binding protein NCS-1
XLOC_028731	tmem129	E3 ubiquitin-protein ligase TM129
XLOC_029242	EaffTmpM009308	Fatty acid-binding protein
XLOC_029253	EaffTmpS009329	predicted protein
XLOC_029254	EaffTmpM009330	SPT transcription factor family member
XLOC_029264	EaffTmpM009259	hypothetical protein AURANDRAFT 63034
XLOC_029315	EaffTmpM009331	n-formylglutamate amidohydrolase
XLOC_029424	EaffTmpM026628	PREDICTED: nocturnin isoform X1
XLOC_029433	EaffTmpM026623	hypothetical protein
XLOC_029549	EaffTmpM011495	hypothetical protein LEMA P075820.1
XLOC_029594	pbo 4	Na(+)/H(+) exchanger protein 7
XLOC_029663	EaffTmpM028237	F-box/WD repeat-containing protein 1A
XLOC_029883	lbp 5	Fatty acid-binding protein homolog 5
XLOC_030077	EaffTmpS009797	RNA-directed DNA polymerase from mobile element jockey, partial
XLOC_030078	Htatsf1	HIV Tat-specific factor 1 homolog
XLOC_030109	MSI1	RNA-binding protein Musashi homolog 1
XLOC_030121	EaffTmpM009851	GF11443
XLOC_030142	lolal	Longitudinals lacking protein-like
XLOC_030143	lolal	Longitudinals lacking protein-like
XLOC_030172	EaffTmpM009870	tripsin, putative
XLOC_030249	EaffTmpS027608	Actin, clone 403
XLOC_030317	Plscr3	Phospholipid scramblase 3
XLOC_030452	PLA2G4A	Cytosolic phospholipase A2
XLOC_030457	EaffTmpM002315	Innexin inx2
XLOC_030464	EaffTmpM002323	dynein heavy chain
XLOC_030540	Gxylt2	Glucoside xylosyltransferase 2
XLOC_030559	Wdr54	WD repeat-containing protein 54
XLOC_030573	CASKIN1	Caskin-1
XLOC_030864	EaffTmpM012398	HEAT repeat-containing protein 5B
XLOC_031084	Hadhb	Trifunctional enzyme subunit beta, mitochondrial
XLOC_031344	Naa25	N-alpha-acetyltransferase 25, NatB auxiliary subunit
XLOC_031352	SEPSECS	O-phosphoseryl-tRNA(Sec) selenium transferase
XLOC_031358	Srsf5	Serine-arginine protein 55

XLOC_031444	METTL3	putative calcium-binding tyrosine phosphorylation-regulated protein
XLOC_031745	EaffTmpM028439	GG10482
XLOC_031867	EaffTmpM009347	Protein mesh
XLOC_032199	Hnf4	Transcription factor HNF-4 homolog
XLOC_032347	OIT3	Uromodulin
XLOC_032490	Orct	Organic cation transporter protein
XLOC_032499	CML8	Calmodulin-like protein 8
XLOC_032525	CG12034	Putative neutral sphingomyelinase
XLOC_032527	Syngr2	Synaptogyrin-2
XLOC_032544	Ufl1	E3 UFM1-protein ligase 1
XLOC_032598	Inpp5e	72 kDa inositol polyphosphate 5-phosphatase
XLOC_032620	PCSK2	Neuroendocrine convertase 2
XLOC_032668	Tret1 2	Facilitated trehalose transporter Tret1
XLOC_032682	NHA-6	Na+/H+ Antiporter, paralog 6
XLOC_032683	NHA-5	Na+/H+ Antiporter, paralog 5
XLOC_032685	NHA-3	Na+/H+ Antiporter, paralog 3
XLOC_032687	NHA-1-frag	Na+/H+ Antiporter, paralog 1 (fragment)
XLOC_032688	NHA-1	Na+/H+ Antiporter, paralog 1
XLOC_033435	CRAM	Cysteine-rich, acidic integral membrane protein
XLOC_033437	infB	Translation initiation factor IF-2
XLOC_033501	w	Protein white
XLOC_033807	Pka C3	Protein kinase DC2
XLOC_033810	DSCR3	Down syndrome critical region protein 3 homolog
XLOC_033986	EaffTmpM011020	Pecanex-like protein 1
XLOC_034122	EDEM3	ER degradation-enhancing alpha-mannosidase-like protein 3
XLOC_034137	Tpcn1	Two pore calcium channel protein 1
XLOC_034145	inx1	Innexin inx2
XLOC_034153	OPCML	Opioid-binding protein/cell adhesion molecule
XLOC_034164	Prkg1	cGMP-dependent protein kinase 1
XLOC_034263	EaffTmpS009452	hypothetical protein HELRODRAFT 164290
XLOC_034584	rsad2	Radical S-adenosyl methionine domain-containing protein 2
XLOC_034663	EaffTmpM010554	hypothetical protein L798 10568
XLOC_034707	EaffTmpM010552	AGAP004872-PA-like protein

XLOC_034709	EaffTmpM010553	PREDICTED: uncharacterized protein LOC100118488
XLOC_034958	EaffTmpM001953	Thioredoxin H1
XLOC_034959	Slc6a7	Sodium-dependent proline transporter
XLOC_034960	Slc6a5	Sodium- and chloride-dependent glycine transporter 2
XLOC_035022	Ptp10D	Tyrosine-protein phosphatase 10D
XLOC_035110	Ky	Kyphoscoliosis peptidase
XLOC_035138	NELL1	Protein kinase C-binding protein NELL1
XLOC_035139	NELL1	Protein kinase C-binding protein NELL2
XLOC_035160	EaffTmpM002088	putative BR serine/threonine-protein kinase
XLOC_035173	Dagla	Sn1-specific diacylglycerol lipase alpha
XLOC_035251	Slc45a2	Membrane-associated transporter protein
XLOC_035422	EaffTmpS011169	AGAP004872-PA-like protein
XLOC_035432	Osbpl1a	Oxysterol-binding protein-related protein 2
XLOC_035435	COPS6	COP9 signalosome complex subunit 6
XLOC_035467	KCNJ18	Inward rectifier potassium channel 18
XLOC_035585	Nudt1	7,8-dihydro-8-oxoguanine triphosphatase
XLOC_035586	SMAP2	Stromal membrane-associated protein 2
XLOC_035588	AP1G1	AP-1 complex subunit gamma-1
XLOC_035757	gnptab	N-acetylglucosamine-1-phosphotransferase subunits alpha/beta
XLOC_035810	PGAP1	GPI inositol-deacylase
XLOC_035812	PGAP1	GPI inositol-deacylase
XLOC_035851	NKA-b-5	Na+/K+-ATPase, subunit beta, paralog 5
XLOC_035880	EaffTmpM011353	hypothetical protein
XLOC_035882	FN1	PREDICTED: receptor-type tyrosine-protein phosphatase H
XLOC_035960	EaffTmpM013716	hypothetical protein
XLOC_035967	TNPO1	Transportin-1
XLOC_035969	Tnpo2	Transportin-2
XLOC_035986	SAR1A	GTP-binding protein SAR1a
XLOC_035991	EaffTmpM013774	Protein msta, isoform A
XLOC_035992	EaffTmpM013776	ferredoxin
XLOC_036007	EaffTmpM013732	PREDICTED: uncharacterized protein DKFZp434B061-like
XLOC_036013	bgm	Very long-chain-fatty-acid-CoA ligase bubblegum
XLOC_036016	EaffTmpM013752	Compound eye opsin BCRH2

XLOC_036020	NELF B	Negative elongation factor B
XLOC_036139	DHCR24	Delta(24)-sterol reductase
XLOC_036140	DHCR24	Delta(24)-sterol reductase
XLOC_036171	EaffTmpM029216	hypothetical protein LOTGIDRAFT 159314
XLOC_036371	EaffTmpM011192	DOMON domain containing protein
XLOC_036449	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_036450	setd7	Histone-lysine N-methyltransferase SETD7
XLOC_036505	Aldh18a1	Delta-1-pyrroline-5-carboxylate synthase
XLOC_036597	xynB	Small proline-rich protein 3
XLOC_036712	SPRR3	Small proline-rich protein 3
XLOC_036940	EaffTmpM009997	GD20657
XLOC_037059	SMYD4	SET and MYND domain-containing protein 4
XLOC_037063	Svep1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
XLOC_037259	abcG23	ABC transporter G family member 23
XLOC_037388	EaffTmpM011945	PREDICTED: uncharacterized protein DDB G0271670-like isoform X1
XLOC_037547	SSPO	SCO-spondin
XLOC_037554	elovl6	Protein PRQFV-amide
XLOC_037593	YTHDF2	YTH domain-containing family protein 2
XLOC_037601	EaffTmpM012205	hypothetical protein BRAFLDRAFT 63701
XLOC_037708	EaffTmpM012346	hypothetical protein AURANDRAFT 63319
XLOC_037753	NBC	Na+, HCO3- cotransporter
XLOC_037824	Supt20h	Transcription factor SPT20 homolog
XLOC_037826	SINAT1	Putative E3 ubiquitin-protein ligase SINAT1

Table S7. GO term enrichment analysis on list of candidate genes in the saline habitats (Montmagny or L'Isle Verte) detected with F_{ST_Window}. Category: b - biological process, c - cellular component, m - molecular function. Count: number of genes counted with the given function in the list of candidate genes under selection. Total: total number of genes with that GO term in the genome. p-value: p-value obtained with 10,000

permutations. Genes: candidate genes with that given GO term, the number of genes can be larger than "count" because if multiple genes are present in the same region under selection only one is counted.

GO Term	Category	Name	Count	Total	p-value	Genes
GO:0015294	m	solute:cation symporter activity	11	144	0.0008	Slc6a18, SLC2A13, EaffTmpM007749, SLC6A5, MRTO4, EAFF017857, Tret1_2, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, NKCC-frag
GO:0015293	m	symporter activity	11	161	0.0021	Slc6a18, SLC2A13, EaffTmpM007749, SLC6A5, MRTO4, EAFF017857, Tret1_2, Tret1, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, NKCC-frag
GO:0007588	b	excretion	12	189	0.0024	AQP3, ClC_a, EAFF012604, mec_2, MRTO4, mec_2, sto_2, sto_2, ARA1, Kif28p, EAFF026071, slc5a9, 5_HT1B, NKCC-frag
GO:0015291	m	secondary active transmembrane transporter activity	15	283	0.0029	Slc6a18, SLC2A13, EaffTmpM007749, NHA-7, NHA-5, NHA-4, EAFF012604, SFXN1, SLC6A5, MRTO4, EAFF017857, Tret1_2, Tret1, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B, NKCC-frag
GO:0022853	m	active ion transmembrane transporter activity	16	330	0.0031	Slc6a18, VhaAC39_1, SLC2A13, EaffTmpM007749, NHA-7, NHA-5, NHA-4, EAFF012604, SFXN1, SLC6A5, MRTO4, EAFF017857, Tret1_2, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B, NKCC-frag
GO:0006654	b	phosphatidic acid biosynthetic process	4	32	0.0036	PLA2G4A, Gpdh1, Pnpla2, rdgA, rdgA
GO:0015385	m	sodium:hydrogen antiporter activity	4	28	0.0039	NHA-7, NHA-5, NHA-4, EAFF012604, 5_HT1B
GO:0046473	b	phosphatidic acid metabolic process	4	33	0.004	PLA2G4A, Gpdh1, Pnpla2, rdgA, rdgA
GO:1902475	b	L-alpha-amino acid transmembrane transport	7	97	0.0042	Slc6a18, EaffTmpM007749, SFXN1, SLC6A5, MRTO4, EAFF017857, Slc13a2, EAFF025707
GO:0015370	m	solute:sodium symporter activity	8	109	0.0042	Slc6a18, EaffTmpM007749, SLC6A5, MRTO4, Slc6a1,

						SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, NKCC-frag
GO:0015081	m	sodium ion transmembrane transporter activity	16	321	0.0042	Slc6a18, Nach-PPK28, Gpdh1, EaffTmpM007749, NHA-7, NHA-5, NHA-4, EAFF012604, mec_2, SLC6A5, MRTO4, mec_2, sto_2, sto_2, Stard3, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B, NKCC-frag
GO:1903649	b	regulation of cytoplasmic transport	5	52	0.0043	EAFF002277, eipr1, DNAJC13, DNAJC13, DNAJC13, EAFF020112, Kif28p
GO:0015166	m	polyol transmembrane transporter activity	3	14	0.0044	AQP3, SLC2A13, slc5a9
GO:0003333	b	amino acid transmembrane transport	8	122	0.0049	Slc6a18, EaffTmpM007749, SFXN1, SLC6A5, MRTO4, EAFF017857, Slc6a1, SLC6A13, Slc13a2, EAFF025707
GO:0015791	b	polyol transport	3	18	0.0054	AQP3, SLC2A13, slc5a9
GO:0071474	b	cellular hyperosmotic response	4	27	0.0066	Gpdh1, MRTO4, SLC6A13, NKCC-frag
GO:0015171	m	amino acid transmembrane transporter activity	7	106	0.0066	Slc6a18, EaffTmpM007749, SFXN1, SLC6A5, EAFF017857, Slc6a1, SLC6A13, Slc13a2, EAFF025707
GO:0005451	m	monovalent cation:hydrogen antiporter activity	4	31	0.0068	NHA-7, NHA-5, NHA-4, EAFF012604, 5_HT1B
GO:0035725	b	sodium ion transmembrane transport	16	340	0.007	Slc6a18, Nach-PPK28, Gpdh1, EaffTmpM007749, NHA-7, NHA-5, NHA-4, EAFF012604, mec_2, SLC6A5, MRTO4, mec_2, sto_2, sto_2, Stard3, Slc6a1, SLC6A13, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B, NKCC-frag
GO:0051139	m	metal ion:proton antiporter activity	4	34	0.0074	NHA-7, NHA-5, NHA-4, EAFF012604, 5_HT1B
GO:0005283	m	sodium:amino acid symporter activity	4	36	0.0087	Slc6a18, EaffTmpM007749, SLC6A5, Slc6a1, SLC6A13
GO:0005343	m	organic acid:sodium symporter activity	5	59	0.009	Slc6a18, EaffTmpM007749, SLC6A5, Slc6a1, SLC6A13, Slc13a3, Slc13a2, nac_1, EAFF025707
GO:0015179	m	L-amino acid transmembrane transporter activity	6	89	0.0091	Slc6a18, EaffTmpM007749, SFXN1, SLC6A5, EAFF017857, Slc13a2, EAFF025707
GO:0015643	m	toxic substance binding	3	18	0.0094	EAFF002295, ACAD10, CHRNA7

Table S8. GO term enrichment analysis on list of candidate genes in the saline habitats (Montmagny or L'Isle Verte) detected with F_{ST_MaxSNP}. Category: b - biological process, c - cellular component, m - molecular function. Count: number of genes counted with the given function in the list of candidate genes under selection. Total: total number of genes with that GO term in the genome. p-value: p-value obtained with 10,000 permutations. Genes: candidate genes with that given GO term, the number of genes can be larger than "count" because if multiple genes are present in the same region under selection only one is counted.

GO Term	Category	Name	Count	Total	p-value	Genes
GO:0090533	c	cation-transporting ATPase complex	5	15	0.0001	NHE-X-c, EAFF012604, Ca_P60A, NKA-a-5, NKA-a-2
GO:0098533	c	ATPase dependent transmembrane transport complex	5	15	0.0001	NHE-X-c, EAFF012604, Ca_P60A, NKA-a-5, NKA-a-2
GO:0005402	m	cation:sugar symporter activity	5	18	0.0003	SLC2A13, At3g05155, Tret1_2, Slc2a1, slc5a9
GO:1902600	b	proton transmembrane transport	15	163	0.0003	VhaAC39_1, SLC2A13, pbo_4, NHE-X-c, NHA-7, NHA-4, NBC, EAFF012604, At3g05155, Ca_P60A, Tret1_2, NKA-a-5, Slc2a1, NKA-a-2, SLC4A8, 5_HT1B
GO:0005351	m	sugar:hydrogen symporter activity	4	11	0.0004	SLC2A13, At3g05155, Tret1_2, Slc2a1
GO:0002026	b	regulation of the force of heart contraction	9	61	0.0004	Mlc1, Mhc, Mhc, NHE-X-c, EAFF012604, Ca_P60A, NKA-a-5, PKC1, NKA-a-2
GO:0015149	m	hexose transmembrane transporter activity	5	24	0.0015	SLC2A13, At3g05155, Slc2a1, slc5a9, Tret1
GO:1903279	b	regulation of calcium:sodium antiporter activity	4	15	0.0018	NHE-X-c, EAFF012604, NKA-a-5, NKA-a-2

GO:0015078	m	hydrogen ion transmembrane transporter activity	11	129	0.0018	VhaAC39_1, SLC2A13, pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, At3g05155, Ca_P60A, Tret1_2, Slc2a1, 5_HT1B
GO:0015385	m	sodium:hydrogen antiporter activity	5	28	0.0019	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, 5_HT1B
GO:0015145	m	monosaccharide transmembrane transporter activity	5	27	0.0028	SLC2A13, At3g05155, Slc2a1, slc5a9, Tret1
GO:0051119	m	sugar transmembrane transporter activity	5	27	0.0028	SLC2A13, At3g05155, Slc2a1, slc5a9, Tret1
GO:0005451	m	monovalent cation:hydrogen antiporter activity	5	31	0.0029	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, 5_HT1B
GO:0051139	m	metal ion:proton antiporter activity	5	34	0.0029	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, 5_HT1B
GO:0045989	b	positive regulation of striated muscle contraction	6	38	0.0029	pbo_4, MYLK, EAFF012604, Ca_P60A, NKA-a-5, NKA-a-2
GO:0022853	m	active ion transmembrane transporter activity	20	330	0.0033	Slc6a18, VhaAC39_1, SLC2A13, pbo_4, NHE-X-c, NHA-7, NHA-4, NBC, EAFF012604, SFXN1, SLC6A5, At3g05155, Ca_P60A, Tret1_2, NKA-a-5, Slc2a1, Slc6a1, SLC6A13, NKA-a-2, SLC4A8, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B
GO:0015299	m	solute:hydrogen antiporter activity	5	37	0.0037	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, 5_HT1B
GO:0005355	m	glucose transmembrane transporter activity	4	18	0.0038	At3g05155, Slc2a1, slc5a9, Tret1
GO:0015291	m	secondary active transmembrane transporter activity	18	283	0.0043	Slc6a18, SLC2A13, pbo_4, NHE-X-c, NHA-7, NHA-4, NBC, EAFF012604, SFXN1, SLC6A5, At3g05155, Tret1_2, Tret1, NKA-a-5, Slc2a1, Slc6a1, SLC6A13, NKA-a-2, SLC4A8, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707, 5_HT1B
GO:0015294	m	solute:cation symporter activity	11	144	0.0058	Slc6a18, SLC2A13, NBC, SLC6A5, At3g05155, Tret1_2, Slc2a1, Slc6a1, SLC6A13, SLC4A8, slc5a9, Slc13a3, Slc13a2, nac_1, EAFF025707
GO:0015491	m	cation:cation antiporter activity	7	64	0.0068	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, NKA-a-5, NKA-a-2, 5_HT1B
GO:0015298	m	solute:cation antiporter activity	7	66	0.0069	pbo_4, NHE-X-c, NHA-7, NHA-4, EAFF012604, NKA-a-5, NKA-a-2,

						5_HT1B
GO:2000649	b	regulation of sodium ion transmembrane transporter activity	9	94	0.0073	Gpdh1, NHE-X-c, EAFF012604, mec_2, mec_2, sto_2, sto_2, Stard3, NKA-a-5, NKA-a-2, 5_HT1B
GO:0003010	b	voluntary skeletal muscle contraction	4	19	0.0078	Mhc, Mhc, MYLK, Ca_P60A
GO:0014721	b	twitch skeletal muscle contraction	4	19	0.0078	Mhc, Mhc, MYLK, Ca_P60A
GO:0014724	b	regulation of twitch skeletal muscle contraction	4	19	0.0078	Mhc, Mhc, MYLK, Ca_P60A
GO:0015297	m	antiporter activity	9	113	0.0095	pbo_4, NHE-X-c, NHA-7, NHA-4, NBC, EAFF012604, NKA-a-5, NKA-a-2, SLC4A8, 5_HT1B
GO:0000146	m	microfilament motor activity	5	54	0.0096	Mhc, Mhc, MYO5A, YAP1, MYO18A

Table S9. GO term enrichment analysis on list of candidate genes in Lake Ontario detected with $PBE_{\_MaxSNP}$. Category: b - biological process, c - cellular component, m - molecular function. Count: number of genes counted with the given function in the list of candidate genes under selection. Total: total number of genes with that GO term in the genome. p-value: p-value obtained with 10,000 permutations. Genes: candidate genes with that given GO term, the number of genes can be larger than "count" because if multiple genes are present in the same region under selection only one is counted.

GO Term	Category	Name	Count	Total	p-value	Genes
GO:0006842	b	tricarboxylic acid transport	5	17	0.0003	Indy, SFXN1, SLC13A5, Slc13a2, Slc13a3
GO:0015740	b	C4-dicarboxylate transport	7	36	0.0003	slc25a40, Indy, EaffTmpM007749, SLC13A5, slc38a7, Slc13a2, Slc13a3

GO:0045124	b	regulation of bone resorption	5	24	0.0013	CA-14, Pdk1, Nf1, PKC1, PKC1
GO:0002026	b	regulation of the force of heart contraction	9	61	0.0013	Mlc1, SLC8A1, Mhc, EAFF012604, PKC1, PKC1, NKA-a-2, NHE2_5, NHE2_5, NKA-a-1
GO:0005697	c	telomerase holoenzyme complex	4	20	0.0031	HNRNPUL1, Hnrnpul1, NHP2, snrpd3
GO:0061064	b	negative regulation of nematode larval development	5	31	0.0045	PCSK2, Prkg1, EAFF010208, NPR2, hdac3
GO:0015291	m	secondary active transmembrane transporter activity	22	283	0.0046	Slc6a18, Indy, SLC8A1, Slc20a1, SLC2A13, EaffTmpM007749, Orct, NHA-7, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, EAFF012604, SFXN1, SLC13A5, Tret1, SLC4A10, SLC22A5, Slc2a1, Slc6a1, NKA-a-2, NHE2_5, NHE2_5, NKA-a-1, Slc13a2, Slc13a3
GO:0003351	b	epithelial cilium movement involved in extracellular fluid movement	6	46	0.0048	Dnah5, EAFF020116, DNAH3, DNAH3, EAFF026038, EAFF026055, EAFF026928, KIF21B, Kif21a
GO:0097381	c	photoreceptor disc membrane	5	25	0.0056	ATP11A, NPR2, EAFF012601, PKC1, PKC1
GO:0090533	c	cation-transporting ATPase complex	4	15	0.0063	EAFF012604, NKA-a-2, NHE2_5, NKA-a-1
GO:0098533	c	ATPase dependent transmembrane transport complex	4	15	0.0063	EAFF012604, NKA-a-2, NHE2_5, NKA-a-1
GO:0070778	b	L-aspartate transmembrane transport	4	19	0.0063	EaffTmpM007749, SLC13A5, slc38a7, Slc13a2
GO:0061067	b	negative regulation of dauer larval development	4	19	0.0067	PCSK2, Prkg1, EAFF010208, NPR2
GO:0010911	b	regulation of isomerase activity	3	10	0.0074	HNRNPUL1, Hnrnpul1, Plscr1
GO:0010912	b	positive regulation of isomerase activity	3	10	0.0074	HNRNPUL1, Hnrnpul1, Plscr1
GO:2000371	b	regulation of DNA topoisomerase (ATP-hydrolyzing) activity	3	10	0.0074	HNRNPUL1, Hnrnpul1, Plscr1
GO:2000373	b	positive regulation of DNA topoisomerase (ATP-hydrolyzing) activity	3	10	0.0074	HNRNPUL1, Hnrnpul1, Plscr1
GO:0046850	b	regulation of bone remodeling	5	33	0.0078	CA-14, Pdk1, Nf1, PKC1, PKC1

GO:0017153	m	sodium:dicarboxylate symporter activity	3	13	0.008	SLC13A5, Slc13a2, Slc13a3
GO:0006637	b	acyl-CoA metabolic process	9	99	0.0088	gpat3, sucg_1, bgm, acsbg2, acsbg2, ACSBG2, ACSBG1, ACSBG2, SLC27A4, AAEL011789, HSD17B4, Pdk1, Acsl4, Acsl3
GO:0035383	b	thioester metabolic process	9	99	0.0088	gpat3, sucg_1, bgm, acsbg2, acsbg2, ACSBG2, ACSBG1, ACSBG2, SLC27A4, AAEL011789, HSD17B4, Pdk1, Acsl4, Acsl3
GO:0022853	m	active ion transmembrane transporter activity	23	330	0.0096	Slc6a18, Indy, SLC8A1, surf1, Slc20a1, VhaAC39_1, SLC2A13, EaffTmpM007749, Orct, NHA-7, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, EAFF012604, SFXN1, SLC13A5, SLC4A10, SLC22A5, Slc2a1, Slc6a1, NKA-a-2, NHE2_5, NHE2_5, NKA-a-1, Slc13a2, Slc13a3

Table S10. GO term enrichment analysis on list of candidate genes in Lake Michigan detected with PBE_{Window}. Category: b - biological process, c - cellular component, m - molecular function. Count: number of genes counted with the given function in the list of candidate genes under selection. Total: total number of genes with that GO term in the genome. p-value: p-value obtained with 10,000 permutations. Genes: candidate genes with that given GO term, the number of genes can be larger than "count" because if multiple genes are present in the same region under selection only one is counted.

GO Term	Category	Name	Count	Total	p-value	Genes
GO:0030007	b	cellular potassium ion homeostasis	6	33	0.0007	slo, EAFF008536, slo, KCNJ18, NKA-a-5, NKA-a-2, KCNJ2, NKA-a-1

GO:1990573	b	potassium ion import across plasma membrane	6	32	0.0019	KCNJ18, SLC12A6, SLC12A6, NKA-a-5, NKA-a-2, KCNJ2, NKA-a-1
GO:0086037	m	P-type sodium:potassium-exchanging transporter activity involved in regulation of cardiac muscle cell membrane potential	3	7	0.0036	NKA-a-5, NKA-a-2, NKA-a-1
GO:0099520	m	ion antiporter activity involved in regulation of presynaptic membrane potential	3	7	0.0036	NKA-a-5, NKA-a-2, NKA-a-1
GO:0055075	b	potassium ion homeostasis	7	55	0.0036	slo, EAFF008536, slo, KCNJ18, SLC12A6, NKA-a-5, NKA-a-2, KCNJ2, NKA-a-1
GO:0005391	m	sodium:potassium-exchanging ATPase activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0008554	m	P-type sodium transporter activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0008556	m	potassium-transporting ATPase activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0005890	c	sodium:potassium-exchanging ATPase complex	3	8	0.0045	NKA-a-5, NKA-a-2, NKA-a-1
GO:0010248	b	establishment or maintenance of transmembrane electrochemical gradient	4	19	0.0047	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0099587	b	inorganic ion import across plasma membrane	10	86	0.0058	pbo_4, KCNJ18, SLC12A6, SLC12A6, Slc9a9, NKA-a-5, nAChRbeta2, NKA-a-2, SLC26A11, KCNJ2, NKA-a-1
GO:0060075	b	regulation of resting membrane potential	4	18	0.0066	KCNJ18, NKA-a-5, NKA-a-2, KCNJ2
GO:0098739	b	import across plasma membrane	15	171	0.0074	Slc6a7, EAFF003481, Akt1, EaffTmpM007749, pbo_4, KCNJ18, SLC12A6, SLC12A6, Slc9a9, SLC27A1, NKA-a-5, nAChRbeta2, NKA-a-2, SLC26A11, KCNJ2, NKA-a-1
GO:0070634	b	transepithelial ammonium transport	3	11	0.0094	SLC12A6, rhbg, NKCC2

Table S11. GO term enrichment analysis on list of candidate genes in Lake Michigan detected with PBE$_{\text{MaxSNP}}$. Category: b - biological process, c -

cellular component, m - molecular function. Count: number of genes counted with the given function in the list of candidate genes under selection. Total: total number of genes with that GO term in the genome. p-value: p-value obtained with 10,000 permutations. Genes: candidate genes with that given GO term, the number of genes can be larger than "count" because if multiple genes are present in the same region under selection only one is counted.

GO Term	Category	Name	Count	Total	p-value	Genes
GO:0008097	m	5S rRNA binding	4	13	0.0016	Rpl3, RpL4, Rpl3, ISCW009002
GO:0086037	m	P-type sodium:potassium-exchanging transporter activity involved in regulation of cardiac muscle cell membrane potential	3	7	0.0017	NKA-a-5, NKA-a-2, NKA-a-1
GO:0099520	m	ion antiporter activity involved in regulation of presynaptic membrane potential	3	7	0.0017	NKA-a-5, NKA-a-2, NKA-a-1
GO:0030004	b	cellular monovalent inorganic cation homeostasis	15	153	0.0017	VhaAC39_1, SLC4A11, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, KCNJ18, NBC, CA-14, Slc9a9, NKA-a-5, NKA-a-2, CCAP_R, KCNJ2, NKA-a-1, Mafg
GO:1901570	b	fatty acid derivative biosynthetic process	8	68	0.0018	Acsl1, w, bgm, acsbg2, GCDH, Acsl3, SKIV2L, EAFF027298, Ggt5
GO:0005890	c	sodium:potassium-exchanging ATPase complex	3	8	0.0023	NKA-a-5, NKA-a-2, NKA-a-1
GO:0031125	b	rRNA 3'-end processing	4	16	0.0023	ISCW009002, Exosc3, SKIV2L, REXO1L1P
GO:0055067	b	monovalent inorganic cation homeostasis	18	215	0.0027	VhaAC39_1, SLC4A11, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, EAFF009605, KCNJ18, NBC, EAFF010944, CA-14, Slc9a9, NKA-a-5, NKA-a-2, CCAP_R, rhbg, KCNJ2,

						NKA-a-1, Mafg
GO:0015672	b	monovalent inorganic cation transport	5	26	0.0032	NKA-b-4, NKA-b-5, Pkd1l2, rhbg, Gid4
GO:0036376	b	sodium ion export across plasma membrane	4	15	0.0041	NKA-a-5, NKA-a-2, NKA-a-1, EAFF029248
GO:0005391	m	sodium:potassium-exchanging ATPase activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0008554	m	P-type sodium transporter activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0008556	m	potassium-transporting ATPase activity	4	18	0.0042	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0022853	m	active ion transmembrane transporter activity	24	330	0.0043	Slc6a7, Slc6a5, NKA-b-4, Slc13a3, VhaAC39_1, EAFF008893, SLC4A11, EaffTmpM007749, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, w, EAFF009605, EAFF009606, Slc45a2, NBC, SFXN1, Orct, Slc9a9, SLC13A5, NKA-a-5, Slc2a1, NKA-a-2, Nckx30C, NKA-a-1, Gid4, Slc36a2, Orct
GO:0010248	b	establishment or maintenance of transmembrane electrochemical gradient	4	19	0.0045	NKA-b-4, NKA-a-5, NKA-a-2, NKA-a-1
GO:0001504	b	neurotransmitter uptake	9	84	0.0046	Slc6a7, Slc6a5, MMP14, EaffTmpM007749, w, EAFF009605, EAFF009606, Orct, NKA-a-5, NKA-a-2, eIF3_S8, NKA-a-1
GO:0030007	b	cellular potassium ion homeostasis	5	33	0.0047	KCNJ18, NKA-a-5, NKA-a-2, KCNJ2, NKA-a-1
GO:1902600	b	proton transmembrane transport	14	163	0.0047	EAFF005216, VhaAC39_1, EAFF008893, SLC4A11, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, Slc45a2, NBC, Slc9a9, NKA-a-5, Slc2a1, NKA-a-2, NKA-a-1, Slc36a2
GO:0015179	m	L-amino acid transmembrane transporter activity	9	89	0.005	Slc6a7, Slc6a5, Slc13a3, EaffTmpM007749, w, EAFF009605, EAFF009606, SFXN1, EAFF012638, SLC13A5, slc38a7, Slc36a2
GO:0015807	b	L-amino acid transport	10	106	0.0054	Slc6a7, Slc6a5, Slc13a3, EaffTmpM007749, w, EAFF009605, EAFF009606, SFXN1, EAFF012638,

						SLC13A5, slc38a7, Gid4, Slc36a2
GO:0070778	b	L-aspartate transmembrane transport	4	19	0.0056	Slc13a3, EaffTmpM007749, SLC13A5, slc38a7
GO:0060075	b	regulation of resting membrane potential	4	18	0.0057	KCNJ18, NKA-a-5, NKA-a-2, KCNJ2
GO:0015491	m	cation:cation antiporter activity	8	64	0.0068	pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, Orct, Slc9a9, NKA-a-5, NKA-a-2, Nckx30C
GO:0015298	m	solute:cation antiporter activity	8	66	0.0071	pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, Orct, Slc9a9, NKA-a-5, NKA-a-2, Nckx30C
GO:0015297	m	antiporter activity	11	113	0.0075	SLC4A11, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, NBC, Orct, Slc9a9, NKA-a-5, NKA-a-2, Nckx30C, NKA-a-1
GO:0015291	m	secondary active transmembrane transporter activity	21	283	0.0078	Slc6a7, Slc6a5, Slc13a3, SLC4A11, EaffTmpM007749, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, Slc45a2, NBC, SFXN1, Orct, Slc9a9, SLC13A5, NKA-a-5, Slc2a1, NKA-a-2, Nckx30C, NKA-a-1, Gid4, Slc36a2, Orct
GO:0045989	b	positive regulation of striated muscle contraction	6	38	0.0083	pbo_4, MYLK, NKA-a-5, NKA-a-2, MYLK, NKA-a-1
GO:1902475	b	L-alpha-amino acid transmembrane transport	9	97	0.0088	Slc6a7, Slc6a5, Slc13a3, EaffTmpM007749, w, EAFF009605, EAFF009606, SFXN1, EAFF012638, SLC13A5, slc38a7, Slc36a2
GO:0002726	b	positive regulation of T cell cytokine production	3	12	0.0091	EAFF005509, gata3, rsad2
GO:0006885	b	regulation of pH	11	118	0.0096	VhaAC39_1, SLC4A11, pbo_4, NHA-6, NHA-5, EAFF008247, NHA-3, NHA-1, NBC, CA-14, Slc9a9, CCAP_R, rhbg, Mafg
GO:1990573	b	potassium ion import across plasma membrane	5	32	0.0097	KCNJ18, NKA-a-5, NKA-a-2, KCNJ2, NKA-a-1

Table S12. Outlier SNPs in each freshwater lake within the genomic region of Scaffold 68 containing seven tandem repeats of the NHA gene.

"Position" refers to the genomic coordinates in the i5K *E. carolleeae* reference genome, located at the i5K workspace (https://i5k.nal.usda.gov/Eurytemora_affinis).

Position	Lake Ontario	Lake Michigan	Genomic region	Gene
534149		X	Intron	PCSK2
534322	X		Intron	PCSK2
536985	X		Intron	PCSK2
561936	X	X	Intergenic	NHA-7 - NHA-6
562099		X	Intergenic	NHA-7 - NHA-6
565615		X	Intergenic	NHA-7 - NHA-6
565628	X		Intergenic	NHA-7 - NHA-6
565712		X	Intergenic	NHA-7 - NHA-6
566924	X		Intergenic	NHA-7 - NHA-6
566926	X		Intergenic	NHA-7 - NHA-6
567091		X	Intergenic	NHA-7 - NHA-6
567114		X	Intergenic	NHA-7 - NHA-6
567938	X		Intergenic	NHA-7 - NHA-6
567953	X		Intergenic	NHA-7 - NHA-6
568010	X	X	Intergenic	NHA-7 - NHA-6
568827	X		Intergenic	NHA-7 - NHA-6
579255		X	Exon	NHA-6
580509	X		Exon	NHA-6
581289		X	Exon	NHA-6
586308	X		Intergenic	NHA-6 - NHA-5
586311		X	Intergenic	NHA-6 - NHA-5
586351	X	X	Intergenic	NHA-6 - NHA-5
586413		X	Intergenic	NHA-6 - NHA-5
586437	X	X	Intergenic	NHA-6 - NHA-5
586477		X	Intergenic	NHA-6 - NHA-5
586694		X	Intergenic	NHA-6 - NHA-5
586700		X	Intergenic	NHA-6 - NHA-5

590889		X	Exon	NHA-5
590931	X		Exon	NHA-5
593038	X		Intron	NHA-5
593323		X	Intron	NHA-5
593324		X	Intron	NHA-5
622703	X		Intergenic	NHA-4 - NHA-3
624744		X	Intergenic	NHA-4 - NHA-3
624753		X	Intergenic	NHA-4 - NHA-3
624767		X	Intergenic	NHA-4 - NHA-3
636161		X	Intron	NHA-3
636388		X	Intron	NHA-3
636551		X	Intron	NHA-3
644702	X		Intergenic	NHA-3 - NHA-2
648304		X	Intergenic	NHA-3 - NHA-2
657560		X	Intergenic	NHA-2 - NHA-1
662722	X		Intergenic	NHA-2 - NHA-1
663175		X	Intergenic	NHA-2 - NHA-1
663176		X	Intergenic	NHA-2 - NHA-1
663185		X	Intergenic	NHA-2 - NHA-1
666245	X		Intergenic	NHA-2 - NHA-1
666472		X	Intergenic	NHA-2 - NHA-1
671662	X		Intron	NHA-1
674159		X	Intron	NHA-1
674223	X		Intron	NHA-1

Chapter 6: General Discussion

Understanding how populations adapt to their environments has been a fundamental question in evolutionary biology from its onset (Darwin 1859), and it is ever more pressing in the face of human-induced rapid global environment changes. Although many advances have been made in our understanding of natural selection, many questions regarding the genetic basis of adaptive evolution still remain unanswered. In this thesis, we used genomic and computational approaches to contribute to the body of work surrounding these questions. We used computer simulations to investigate our ability to detect different kinds of selective sweeps, we used genomic and experimental data from a model organism to study the genetic basis of adaptive traits, and we used genomewide information from natural populations of a non-model organism to investigate biological functions under selection. Herein, I discuss the outcome of these projects.

An important step in the study of the genetic basis of adaptation is being able to detect genomic signatures of selection. In the second chapter of this thesis, we compared the power of window-wide approaches and maximum SNP F_{ST} (F_{ST_MaxSNP}) to detect different kinds of selective sweeps and applied them to empirical data of natural populations of *Drosophila melanogaster*. We found that F_{ST_MaxSNP} outperformed the other methods when detecting complete or nearly complete soft sweeps, and underperformed when detecting partial harder sweeps that only reached moderate final

frequencies. When applied to empirical data of *D. melanogaster*, we found that both SNP-level and window-wide F_{ST} were enriched in empirical data based on neutral demographic simulations, but they detected mostly unique regions of the genome and biological functions. Based on the complementarity F_{ST_MaxSNP} showed to window-wide approaches, we employed both strategies when performing genome scans in the following chapters. Ultimately, it would be interesting to see how the inclusion of F_{ST_MaxSNP} as a summary statistic could affect methods that combine multiple sources of evidence to detect signatures of selection, such as approximate Bayesian computation and machine learning (e.g. Schrider & Kern 2016, Sheehan & Song 2016). We acknowledge that in this study we only modeled two populations with relatively simple demographic histories. Therefore, future research could investigate whether similar conclusions will hold true in the face of more complex scenarios.

In the third and fourth chapters, we focused on the genetic basis of adaptive traits in *D. melanogaster*. There is still debate regarding the distribution of allelic effect sizes underlying adaptive traits, as well as the importance of genetic interactions. In the third chapter of this thesis, we used two new panels of recombinant inbred lines (RILs) to identify loci underlying multiple adaptive traits and search for interactions. Complex traits have a polygenic nature, and polygenic adaptation has been proposed to involve many loci of negligible effect sizes (Pritchard *et al*. 2010,

Rockman 2012). Nonetheless, we showed that several complex adaptive traits often have loci of detectable size. Regarding the role of genetic interactions, we did not find evidence of strong pervasive gene-by-gene interaction, but for pigmentation, the trait that we investigated at two temperatures, we found that the environment affected the magnitude of the effect size of adaptive loci. Given the sample size of our analyses, we cannot rule out moderate or weak epistatic interactions among adaptive loci. Additionally, we had to limit epistasis tests to loci with significant additive effects. Despite advances in DNA sequencing, the lack of high throughput phenotyping, as well as the effort required to generate and maintain a large number of inbred lines, limited our power to detect epistasis.

Another current debate regarding the genetic basis of adaptation concerns the number of loci involved as well as the nature of the selective sweeps. In the fourth chapter, we focused on a single population to study the genetic basis of melanism, an adaptive trait in high elevations in sub-Saharan Africa (Bastide *et al*. 2014). We generated twenty-one mapping crosses between dark strains from an adapted population and light strains from a population within the ancestral range of the species. We used these mapping crosses to identify quantitative trait loci (QTLs) underlying three pigmentation traits. A subset of the crosses was also measured at a colder temperature to investigate the effect of the environment on the genetic

basis of the traits. Congruent with the previous chapter and a previous study (Bastide *et al*. 2016), we showed that adaptive melanism has a variable genetic architecture with many loci of weak to strong effect sizes and temperature affects the magnitude of effect sizes. We found that both dark and light parental strains influenced which QTLs were detected in each mapping cross, supporting partial soft sweeps from standing variation as the mechanism underlying adaptive evolution in this case. An expanded version of this work at 15 °C would have been ideal, given that this is closer to the condition of the ancestral population, but raising flies at that condition is impractical for the scale we analyzed here. Ultimately, it would also be desirable to investigate the RNA expression patterns of adaptive genes in the different strains studied here to validate our findings. In hindsight, given the results of this chapter, in particular the amount of adaptive variant that still seems to exist in the ancestral Zambian population, an experimental design that involved more crosses with different Zambian inbred lines, such as five Zambian and five Ethiopian, would have been ideal to explore the degree to which standing genetic variation contributed to pigmentation adaptation. Overall, we showed that adaptive melanism has a polygenic basis with several loci of moderate to strong effect and that multiple paths to similar outcomes can co-exist in the same population. Lastly, it is worth noting that, although with smaller samples sizes, similar results were found in other adaptive

quantitative traits in *D. melanogaster*, such as ethanol resistance (Sprengelmeyer *et al*. 2021) and thorax and wing length (Sprengelmeyer *et al*. 2022), suggesting that our conclusions are not limited to high altitude melanism.

In chapter five, instead of starting from the adaptive trait we started from the genome to find candidate biological functions underlying adaptation to a novel environment. We compared recent adaptation to freshwater in invasive populations of the copepod *Eurytemora affinis* complex to local adaptation to different salinities in the last ~17 ky. We found that ion transport-related genes were enriched in both instances and have likely played a key role in adaptation at both timescales and through the salinity gradient. However, at the older timescale, we also found enrichment for genes related to the regulation of ion transport, suggesting that a more fine-tuned control of ion transport might be a later step in adaptation process. A demographic model of the history of these populations could further assist in identifying candidate genes under selection. Other methods to detect signatures of selection also exist, such as BayPass (Gautier 2015) and GRoSS (Refoyo-Martínez *et al*. 2019), and could be used to investigate whether similar results would be found. *Eurytemora affinis* is distributed throughout the Northern Hemisphere and has independently invaded freshwater habitats several times (Lee 1999); replication of this study on other populations could also offer insights into

the generality of our conclusions. Lastly, it is important to keep in mind that other factors besides salinity also vary among these populations, including food sources and the pathogens present in the environment. Although ion transport was the most striking biological function under selection, it is likely that other functions might have contributed to local adaptation among these populations as well.

Taken together, the results of this thesis suggest that the use of a broad range of population genetics statistics can increase the likelihood of detecting genetic variants under selection. However, given the degree to which adaptive variants might be common prior to selection and that many adaptive variants don't reach fixation, it is worth keeping in mind that we might still be missing many important targets of selection. Additional knowledge about the systems being studied, as well as complementary experiments, could be valuable in the task of identifying candidate genes under selection.

We hope that our methodological contributions may assist evolutionary biologists in identifying selective sweeps and that our insights regarding adaptive traits and adaptation to novel environments contribute to paving the way toward a fuller understanding of the adaptive process.

References

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., & Pool, J. E. (2016). A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics*, 204(3), 1307–1319. https://doi.org/10.1534/genetics.116.192492.

Bastide, H., Yassin, A., Johanning, E. J., & Pool, J. E. (2014). Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan Africa. *BMC Evolutionary Biology*, 14(1), 179. https://doi.org/10.1186/s12862-014-0179-y.

Darwin, C. (1859). *On the origin of species*.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, 201(4), 1555-1579. https://doi.org/10.1534/genetics.115.181453.

Lee, C. E. (1999). Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, 53(5), 1423–1434. https://doi.org/10.2307/2640889.

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation–not by sweeps alone. *Nature Reviews Genetics*, 11(10), 665-667. https://doi.org/10.1038/nrg2880.

Refoyo-Martínez, A., da Fonseca, R. R., Halldórsdóttir, K., Árnason, E., Mailund, T., & Racimo, F. (2019). Identifying loci under positive selection in complex population histories. *Genome Research*, 29(9), 1506-1520. https://doi.org/10.1101/gr.246777.118.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 66(1), 1-17. https://doi.org/10.1111/j.1558-5646.2011.01486.x.

Schrider, D. R., & Kern, A. D. (2016). S/HIC: Robust identification of soft and hard sweeps using machine learning. *PLoS Genetics*. 12(3):e1005928. doi:10.1371/journal.pgen.1005928.

Sheehan, S., & Song, Y. S. (2016). Deep learning for population genetic inference. *PLoS Computational Biology*. 12(3):e1004845. doi:10.1371/journal.pcbi.1004845.

Sprengelmeyer, Q. D., Lack, J. B., Braun, D. T., Monette, M. J., & Pool, J. E. (2022). The evolution of larger size in high-altitude *Drosophila melanogaster* has a variable genetic architecture. *G3*, 12(3), jkab454. https://doi.org/10.1093/g3journal/jkab454.

Sprengelmeyer, Q. D., & Pool, J. E. (2021). Ethanol resistance in *Drosophila melanogaster* has increased in parallel cold-adapted populations and shows a variable genetic architecture within and between populations. *Ecology and Evolution*, 11(21), 15364-15376. https://doi.org/10.1002/ece3.8228