Where's the Money? Inversions, Genes, and the Hunt for Genomic Targets of Selection

Mark Kirkpatrick*,[1] and Andrew Kern†
*Section of Integrative Biology, University of Texas, Austin, Texas 78712 and †Department of Genetics, Rutgers University, Piscataway, New Jersey 08854

In this commentary, Kirkpatrick and Kern discuss Cheng et al., "Ecological genomics of Anopheles gambiae along a latitudinal cline: a population-resequencing approach," which is published in this issue of GENETICS.

Legend has it that, when asked why he robbed banks, Willie Sutton replied, "Because that's where the money is" (Sutton 1976). This impeccable logic is equally useful to guide the hunt for genes that are evolving adaptively. Starting with Dobzhansky's pioneering work, dozens of polymorphic chromosome inversions have been found that show signatures of strong selection: steep clines, seasonal cycles in frequencies, and correlations with environmental variables (Dobzhansky 1937; Krimbas and Powell 1992; Hoffmann and Rieseberg 2008). But despite decades of research on inversions, we still know little about which of the genes they carry are the targets of selection or how the polymorphisms themselves are maintained.

In this issue of GENETICS, Cheng et al. exploit an inversion in the mosquito *Anopheles gambiae* to identify genes that are locally adapted. The species in question, which is the best studied (and most potent) of several malaria vectors in equatorial Africa, segregates for over 80 inversion polymorphisms (Pombi *et al.* 2008). Cheng and coworkers focused most of their attention on *2La*, a large inversion spanning some 22 Mb and ~10% of all coding regions in the genome. In Cameroon, there is a steep cline running ~500 km from the arid north, where the inversion is almost fixed, to the humid south, where it is nearly absent (Coluzzi et al. 1979; Simard *et al.* 2009). White *et al.* (2007) previously surveyed divergence between the two arrangements using microarrays on a sample of mosquitoes from a single village near the middle of that cline. They found more divergence than expected in regions near each of the two breakpoints and suggested that these regions harbor genes that are targets of spatially varying selection that is maintaining the inversion polymorphism.

Cheng *et al.* (2012) revisit this intriguing situation with the most detailed analysis of sequence variation in an inversion polymorphism of any organism, a milestone in the evolutionary genetics of chromosome rearrangements. They extended an experimental design used in a recent study of a famous latitudinal cline in *Drosophila melanogaster* in Australia (Kolaczkowski *et al.* 2011). Cheng *et al.* collected four samples of mosquitoes: inversion homozygotes from the northern end of the cline, standard homozygotes from the southern end of the cline, and both types of homozygotes from a point at the center of the cline. Individual mosquitoes from each sample were pooled, and libraries for sequencing were created from these pools using a technique that is quickly becoming commonplace. Pooled sequencing designs offer an extremely cost-efficient approach to estimating allele frequencies from populations on a genome-wide scale. The approach has been strengthened by recent statistical work that provides pooled sequencing estimators of the population mutation rate, summaries of the site-frequency spectrum, and population differentiation (Futschik and Schlotterer 2010; Kolaczkowski *et al.* 2011). A weakness of pooled sequencing designs is that all information about patterns of linkage disequilibrium is lost.

The first result to emerge from this research comes from patterns of divergence between populations at the two ends of the cline. Divergence between inverted *2La* from the north and the standard arrangement in the south is impressively large ($F_{ST} = 0.25$) and about twice as great as in colinear parts of the genome ($F_{ST} = 0.12$). Second, Cheng *et al.* (2012) find that the frequencies of single nucleotide polymorphisms (SNPs) within each arrangement are more similar along the cline than they are when sampled from the two arrangements in the same village. These patterns make sense once we recall the basic genetic features of inversions

doi: 10.1534/genetics.112.139899
[1]Corresponding author: Section of Integrative Biology, 1 University Station C0930, University of Texas, Austin, TX 78712. E-mail: kirkp@mail.utexas.edu

(Hoffmann and Rieseberg 2008). While recombination plays by the usual rules in homozygotes, it is almost completely blocked in inversion heterozygotes. A low level of recombination between the arrangements (called "genetic flux" by Cheng *et al.*) results from double-crossover events and gene conversion. Neutral variation is expected to diverge between the arrangements much as it does between species that occasionally hybridize (Kirkpatrick 2010). The overall pattern of divergence between the arrangements at *2La* in the Cameroon is therefore what many had anticipated but that had never been seen before at the resolution allowed by sequencing.

How does this picture from *A. gambiae* compare with studies of other inversions? The study of the latitudinal cline in *D. melanogaster* from eastern Australia (Kolaczkowski *et al.* 2011) is the only one that is comparable. Divergence between inverted and standard arrangements is less than half as large for inversion *In(3R)P* in flies than for inversion *2La* in mosquitoes. Cheng *et al.* (2012) suggest that this difference results from the interesting history of *2La*. In a reversal from the role in which inversions are frequently cast as agents of reproductive isolation, inversion *2La*, now found in *A. gambiae*, likely introgressed under strong positive selection following hybridization with the closely related *Anopheles arabensis* (Besansky *et al.* 2003). If so, recombination between the arrangements was impossible before the introgression occurred. A second (and not mutually exclusive) explanation is that the fly inversion is younger [$\sim N_e$ generations old (Andolfatto *et al.* 2001)] than the mosquito inversion [$\sim 2.6\ N_e$ generations (White *et al.* 2007)]. Following the origin of an inversion, neutral divergence between inverted and uninverted arrangements is expected to increase with time until, ultimately, an equilibrium is reached between recombination and drift [Guerrero (2012), but contrary to Cheng *et al.* (2012)].

Cheng *et al.* (2012) next turn to their main question: What are the genetic targets of selection in *2La*? Their strategy begins by measuring divergence in small windows (1 kb in width) along the length of the inverted region. The general picture looks like the cables of a suspension bridge: highest divergence near the breakpoints, dipping to lowest in the center of the inverted region. That pattern is expected because recombination between standard and inverted chromosomes is lowest near the breakpoints (Navarro *et al.* 1997; Andolfatto *et al.* 2001; Guerrero *et al.* 2012). There is, however, an impressive amount of variation around the basic curve. Cheng *et al.* suggest that the variation may result from the action of many SNPs under spatially varying selection. That interpretation should be treated with caution, however, since they did not compare their data against a null model. Models of inversions show that substantial variation in F_{ST} can result from the basic stochastic nature of the coalescent process (Guerrero *et al.* 2012).

A more compelling case for targets of selection comes from examining the sequences in the diverged regions. Many nonsynonymous SNPs are highly differentiated between the two arrangements (as shown in their figure 7), suggesting that these polymorphisms are involved in local adaptation. Furthermore, when genes contained in windows of high divergence are clustered on the basis of gene ontology, many of the pathways identified had also been implicated in the *Drosophila* inversion cline (Kolaczkowski *et al.* (2011). For example, *EGFR*, *TGF-β*, and *EcR* pathway genes are outliers in both studies. This suggests that different groups of insects may often adapt to similar environmental gradients by evolution of shared sets of genes. These parallel results from *Anopheles* and *Drosophila* provide strong evidence that these genes are under selection and underscore the need for further comparative genetic work along common environmental gradients. In sum, three lines of evidence make the case for local adaptation of these genes along the *Anopheles* cline: very high F_{ST}, an enrichment for nonsynonymous polymorphisms among the most differentiated SNPs, and overlap with the set of genes in flies that are also adapting to environmental gradients.

The strategy used by Cheng *et al.* to identify genes under selection begins with finding genomic regions that are highly diverged in the inverted and standard arrangements. Ultimately, this outlier-based approach is expected to miss many of the loci under selection (Teshima *et al.* 2006). An alternative strategy starts with a population genetic model that incorporates drift, migration, and selection and then finds genes under local adaptation and estimates the forces acting on them by fitting the model to the data. This approach offers both more power and the potential to estimate the strengths of selection and other forces, but incurs the cost of requiring detailed assumptions. Recent work on human populations has come a long way toward providing such a model (Nicholson *et al.* 2002; Coop *et al.* 2010; Hancock *et al.* 2011). It is not clear, however, if that type of model is appropriate for large outbreeding populations under strong selection throughout their genomes, as mosquitoes and flies seem to be. There is still much work to be done in developing the statistical machinery for spatially explicit population genetics.

Those population genetic approaches alone, however, cannot answer the fundamental question of how the inversion polymorphism is maintained. For that we need to quantify the evolutionary forces at work and then show with a plausible model that they can explain the cline. That is not a trivial agenda. The cline shows that selection varies in space, and so fitnesses must be estimated at different points along the cline. Furthermore, inversions have been implicated in nonrandom mating (Feder *et al.* 2003; Lowry and Willis 2010), and, if *2La* also has those effects, then they must be estimated as well. One strategy here is to estimate these parameters by fitting an explicit model to data on the frequencies of the inversion along the cline.

Explaining how the inversion polymorphism is currently maintained does not explain how it was first established. A mechanism that is consistent with the data on hand is that a new inversion can spread because it reduces recombination

between loci that are under local adaptation (Kirkpatrick and Barton 2006). This hypothesis predicts that the genic polymorphisms predate the inversion. A prime candidate for this scenario is a *Heliconius* butterfly in which an inversion captured several loci that contribute to a mimetic color pattern (Joron *et al.* 2011; Jones *et al.* 2012). Alternatively, the inversion could have direct fitness effects, for example, by a deletion or changes in gene expression near the breakpoints. Under this hypothesis, adaptive divergence between the rearrangements accumulated after the inversion was established. With appropriate data, these two possibilities might be distinguished.

Much of the recent effort in evolutionary genetics has been devoted to revealing ever more microscopic views of how selection acts on the genome, with the goal of reaching a more mechanistic understanding of adaptation (Barrett and Hoekstra 2011). Classically, this research begins by identifying large QTL, then genes, and finally nucleotides that affect fitness. A reverse strategy, illustrated by Cheng *et al.* (2012), is opened up by next-generation sequencing: divergence between species or populations is exploited to search for genes undergoing adaptation. Finding those genes is the endgame for some geneticists. For others, however, it is only one step toward the more distant goal of understanding how evolution shapes variation within and divergence between species. Among the host of issues that wait to be resolved is whether there are fundamental differences between inversions that are polymorphic within species *vs.* those that are fixed between species, the role that segregation distortion plays in their evolution, whether inversions are typically under- or overdominant, and why there seem to be consistent differences between the rates of genomic rearrangement in autosomes and sex chromosomes. Answers will depend on a new synthesis that weds genomic data, phenotypic data, and mathematical models.

Literature Cited

Andolfatto, P., F. Depaulis, and A. Navarro, 2001 Inversion polymorphisms and nucleotide variability in Drosophila. Genet. Res. 77: 1–8.

Barrett, R. D. H., and H. E. Hoekstra, 2011 Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. 12: 767–780.

Besansky, N. J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern *et al.*, 2003 Semipermeable species boundaries between Anopheles gambiae and Anopheles arabiensis: evidence from multilocus DNA sequence variation. Proc. Natl. Acad. Sci. USA 100: 10818–10823.

Cheng, C., B. J. White, C. Kamdem, K. Mockaitis, C. Costantini *et al.*, 2012 Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. Genetics 190: 1417–1432.

Coluzzi, M., A. Sabatini, V. Petrarca, and M. A. DiDeco, 1979 Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans. R. Soc. Trop. Med. Hyg. 73: 483–497.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.

Dobzhansky, T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York.

Feder, J. L., F. B. Roethele, K. Filchak, J. Niedbalski, and J. Romero-Severson, 2003 Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics 163: 939–953.

Futschik, A., and C. Schlotterer, 2010 The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics 186: 207–218.

Guerrero, R. F., F. Rousset, and M. Kirkpatrick, 2012 Coalescence patterns for chromosomal inversions in divergent populations. Philos. Trans. R. Soc. Lond. B –Biol. Sci. 367: 430–438.

Hancock, A. M., D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin *et al.*, 2011 Adaptations to climate-mediated selective pressures in humans. PLoS Genet. 7: e1001375.

Hoffmann, A. A., and L. H. Rieseberg, 2008 Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst. 39: 21–42.

Jones, R. T., P. A. Salazar, R. H. Ffrench-Constant, C. D. Jiggins, and M. Joron, 2012 Evolution of a mimicry supergene from a multilocus architecture. Proc. Biol. Sci. 279: 316–325.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee *et al.*, 2011 Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477: 203–206.

Kirkpatrick, M., 2010 How and why chromosome inversions evolve. PLoS Biol. 8: e1000501.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation, and speciation. Genetics 173: 419–434.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011 Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187: 245–260.

Krimbas, C. B., and J. R. Powell, 1992 *Drosophila Inversion Polymorphism*, CRC Press, London.

Lowry, D. B., and J. H. Willis, 2010 A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 8: e1000500.

Navarro, A., E. Betran, A. Barbadilla, and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146: 695–709.

Nicholson, G., A. V. Smith, F. Jónsson, Ó. Gústafsson, K. Stefánsson *et al.*, 2002 Assessing population differentiation and isolation from single-nucleotide polymorphism data. J. R. Stat. Soc. Series B Stat. Methodol. 64: 695–715.

Pombi, M., B. Caputo, F. Simard, M. A. Di Deco, M. Coluzzi *et al.*, 2008 Chromosomal plasticity and evolutionary potential in the malaria vector Anopheles gambiae sensu stricto: insights from three decades of rare paracentric inversions. BMC Evol. Biol. 8: 309.

Simard, F., D. Ayala, G. C. Kamdem, M. Pombi, J. Etouna *et al.*, 2009 Ecological niche partitioning between Anopheles gambiae molecular forms in Cameroon: the ecological side of speciation. BMC Ecol. 9: 17.

Sutton, W., 1976 *Where the Money Was*. Viking Press, New York.

Teshima, K. M., G. Coop, and M. Przeworski, 2006 How reliable are empirical genomic scans for selective sweeps? Genome Res. 16: 702–712.

White, B. J., M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo *et al.*, 2007 Localization of candidate regions maintaining a common polymorphic inversion (2La) in Anopheles gambiae. PLoS Genet. 3: 2404–2414.