MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 848–850

NEWS AND VIEWS

PERSPECTIVE

Genomic islands *of* speciation or genomic islands *and* speciation?

THOMAS L. TURNER* and
MATTHEW W. HAHN†

**Evolution, Ecology, and Marine Biology Department, University of California Santa Barbara, Santa Barbara, CA 93106, USA, †Department of Biology and School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA*

Populations of the malaria mosquito, *Anopheles gambiae*, are comprised of at least two reproductively isolated, sympatric populations. In this issue, White *et al.* (2010) use extensive sampling, high-density tiling microarrays, and an updated reference genome to clarify and expand our knowledge of genomic differentiation between these populations. It is now clear that DNA near the centromeres of all three chromosomes are in near-perfect disequilibrium with each other. This is in stark contrast to the remaining 97% of the assembled genome, where fixed differences between populations have not been found, and many polymorphisms are shared. This pattern, coupled with direct evidence of hybridization in nature, supports models of ''mosaic'' speciation, where ongoing hybridization homogenizes variation in most of the genome while loci under strong selection remain in disequilibrium with each other. However, unambiguously demonstrating that selection maintains the association of these pericentric ''speciation islands'' in the face of gene flow is difficult. Low recombination at all three loci complicates the issue, and increases the probability that selection unrelated to the speciation process alters patterns of variation in these loci. Here, we discuss these different scenarios in light of this new data.

Received 7 December 2009; revision accepted 20 December 2009

Populations which are partially reproductively isolated offer opportunities to study the speciation process at its early stages. Populations of the African malaria mosquito, *Anopheles gambiae*, present such a case, and with the publication of White *et al.* 2010 (this issue) these populations have been investigated to an exceptional degree at the genetic level. In this perspective, we consider the implications of this great effort.

Correspondence: Thomas L. Turner, Fax: +1 805-893-4724; E-mail: tturner@lifesci.ucsb.edu

The complexity of relationships among mosquitoes in the *Anopheles gambiae* clade undermines attempts to provide brief summaries of the situation. Indeed, it is likely that these relationships would have escaped our notice if it was not for their medical importance: some mosquitoes in this group are prolific vectors of human malaria in Africa, while others are not. Because of this fact, an uncommon effort to delineate reproductively isolated groups has commenced. These efforts have resulted in seven named species, all of which are closely related, share considerable genetic variation, and are morphologically indistinguishable as adults (White 1974; Coluzzi *et al.* 2002). Hybrids between the primary malaria vector, *A. gambiae sensu stricto*, and other species have been found in nature (at <<1% frequency), further illustrating the close relationships among these taxa (Coluzzi *et al.* 1979; Costantini *et al.* 2009; Simard *et al.* 2009). However, even within the taxon known as *A. gambiae sensu stricto*, there is now indisputable evidence for further evolutionarily significant divisions. This species is split into two 'molecular forms' (an intentionally ambiguous division) called M-form and S-form (della Torre *et al.* 2001, 2005; Lehmann & Diabaté 2008). These forms can only be distinguished at the molecular level, and are sympatric across much of their range, even to the level of resting inside the same houses (della Torre *et al.* 2005; Diabaté *et al.* 2009). As such, they seem to be the spearhead of an ongoing species radiation in the complex, and the lack of geographic separation of such close species makes them a system of great interest.

Efforts to understand the ecological and behavioral differences between M-and S-form mosquitoes is incipient but ongoing. Though no differences between forms or their hybrids have been detected in the lab, a landmark study demonstrated that they mate assortatively in nature by

Fig. 1 *Anapheles gambiae* blood feeding as photographed by J. Gathany (CDC), courtesy of N. Besansky.

genotyping wild-caught females and the sperm they had stored from previous matings (Tripet *et al.* 2001). It was also recently shown that males of each form differentially establish mating aggregations (Diabaté *et al.* 2009), and evidence is mounting that each form may have a different ecological niche (Lehmann & Diabaté 2008; Costantini *et al.* 2009; Lee *et al.* 2009; Simard *et al.* 2009). As M- and S-form individuals in some well studied areas are separated by additional genomic regions of differentiation which are not common to all M and S populations (Lanzaro *et al.* 1998; Slotman *et al.* 2006; Turner & Hahn 2007; White *et al.* this issue), it is unclear if these phenotypes are consistently different between forms throughout Africa.

In contrast to our still-developing understanding of differences at the phenotype level, much is now known about genetic differentiation between forms. The forms were originally defined based on the lack of heterozygotes at a single biallelic marker on the X chromosome, proximal to the centromere (della Torre *et al.* 2001). When M-form and S-form mosquitoes were compared at ~150,000 markers throughout the genome, it was found that this X chromosome genotype is in near-perfect association (i.e. gametic disequilibrium) with a region proximal to the centromere on chromosome 2 (Turner *et al.* 2005). Using this same genotyping platform, but empowered by a more comprehensive genome assembly, White *et al.* (this issue) have now shown that DNA near the centromere of chromosome 3 is also associated with the M/S-form genotype, which means that DNA near centromeres of all three chromosomes in this species are in near-perfect disequilibrium with each other. With the exception of rare hybrid individuals (more about this below), sequenced loci in these pericentromeric regions show fixed DNA differences between forms, with no polymorphisms shared between them. This is in stark contrast to the remaining 97% of the assembled genome, where fixed differences have not yet been found, and many polymorphisms are shared. Though allele frequency differences in the rest of the genome are often adequate to cluster samples based on their M/S designation, this degree of differentiation is slight compared to fixed differentiation seen in the pericentromeric regions.

A pattern of low F_{ST} values across the majority of the genome, coupled with direct evidence of hybridization in nature, would seem to support models of 'mosaic' speciation, where ongoing hybridization homogenizes variation in most of the genome while loci under strong selection remain in disequilibrium with each other (Wu & Ting 2004; Via & West 2008; Nosil *et al.* 2009). In this model the differentiated regions near the centromeres would also harbor the loci that are responsible for assortative mating and ecological differentiation (so called 'speciation genes'). The rest of the genome is either neutral with respect to species differences, or close enough to neutral that gene flow overwhelms selection. New mutations which are positively selected across all genetic backgrounds would strongly promote gene flow, and empirical data on insecticide resistance mutations suggests that this has indeed occurred (Djogbénou *et al.* 2008, Etang *et al.* 2009). We will call this scenario the 'speciation island' hypothesis. Ample room for debate remains, however, as unambiguously demonstrating that selection maintains the association of these pericentric 'speciation islands' in the face of gene flow is difficult. The linkage of these regions to the centromeres complicates the situation, as these pericentromeric regions have low recombination, so that linked selection unrelated to the speciation process alters patterns of variation in these loci.

A reasonable alternative hypothesis—raised by White and colleagues—is that there is actually little gene flow between the M- and S-forms, and that the low level of differentiation observed on chromosome arms is due to segregating ancestral variation. Under this scenario, the hybrids observed in nature would be mostly F_1 individuals who die before reproducing. If F_1 genotypes are effectively sterile due to ecological or behavioural maladaptation, then the amount of 'realized' gene flow could be near zero. The divergent islands could appear different from the rest of the genome because linked selection in these low recombination regions reduces the effective population size, leading to faster sorting of ancestral variation by fixing alternative haplotypes in each form. Here, we refer to this as the 'incidental island' hypothesis. Note that under both scenarios there could be additional regions of the genome which are differentiated but have not yet been found (the limited resolution of microarray platforms can easily miss small regions).

The finding of a third unlinked island leads White *et al.* to express a healthy dose of scepticism towards the speciation island hypothesis. Is it possible that selection can be strong enough to maintain extremely strong associations between regions on three chromosomes, while still allowing enough gene flow at other loci to retard the inexorable differentiation which would result from isolation? Though the findings of obvious hybrids and low genetic differentiation immediately suggest high levels of gene flow, White *et al.* make the case that these have been misleading indicators of introgression.

Luckily, there are opportunities to empirically test the plausibility of these competing hypotheses. White and colleagues found that three of the five individuals with hybrid genotypes were heterozygous in all three islands (likely F_1s), with the final two individuals hybrid at only one island (apparent backcrosses). This suggests that there is indeed some gene flow through F_1 genotypes, though this needs to be quantified in larger samples. Collections of tens of thousands of individuals have found between 0% and 20% heterozygous individuals at the X-linked island (della Torre *et al.* 2005; Caputo *et al.* 2008; Oliveira *et al.* 2008; Costantini *et al.* 2009; Simard *et al.* 2009), and these collections can be used to check the genotype of the additional islands. An even more direct test of these hypotheses would be to perform linkage analysis to associate reproductive isolation with genetic variation. Though this is surely quite challenging under field conditions, recent studies have cataloged some behavioral and ecological differences between forms, making it possible to associate these phenotypes with genotypes in the wild (Lehmann &

Diabaté 2008; Diabaté *et al.* 2009). This could be done by creating and releasing hybrid genotypes, or by using hybrid genotypes found in nature. Though these hybrids are normally rare, a population has been found in Guinea Bissau which has an abnormally high (20%) rate of heterozygous genotypes at the X chromosome island (Oliveira *et al.* 2008). As this is still a much smaller percentage of heterozygotes than expected under Hardy–Weinberg equilibrium, it does not appear that the isolating barriers have completely collapsed in this population. By genotyping individuals in this population at all three islands, and associating these genotypes with swarming behavior, mate-choice in nature (using stored sperm), and larval environment, the direct association between differentiated islands and extrinsic reproductive isolation may be possible.

References

Caputo B, Nwakanma D, Jawara M *et al.* (2008) *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. *Malaria Journal*, **7**, 182.

Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002) A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science*, **298**, 1415–1418.

Coluzzi M, Sabatini A, Petrarca V *et al.* (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **73**, 483–497.

Costantini C, Ayala D, Guelbeogo WM *et al.* (2009) Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecology*, **9**, 16.

della Torre A, Fanello C, Akogbeto M *et al.* (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Molecular Biology*, **10**, 9–18.

della Torre A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae s.s.* molecular forms. *Insect Biochemistry and Molecular Biology*, **35**, 755–769.

Diabaté A, Dao A, Yaro AS *et al.* (2009) Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 4215–4222.

Djogbénou L, Chandre F, Berthomieu A *et al.* (2008) Evidence of introgression of the *ace-1*^{R} mutation and of the *ace-1* duplication in West African *Anopheles gambiae s. s. PLoS One*, **3**, e2172.

Etang J, Vicente JL, Nwane P *et al.* (2009) Polymorphism of intron-1 in the voltage-gated sodium channel gene of *Anopheles gambiae s.s.* populations from Cameroon with emphasis on insecticide knock-down resistance mutations. *Molecular Ecology*, **18**, 3076–3086.

Lanzaro GC, Toure YT, Carnahan J *et al.* (1998) Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14260–14265.

Lee Y, Cornel AJ, Meneses CR *et al.* (2009) Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae sensu stricto*. *Malaria Journal*, **8**, 75.

Lehmann T, Diabaté A (2008) The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infection, Genetics and Evolution*, **8**, 737–746.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Oliveira E, Salgueiro P, Palsson K *et al.* (2008) High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *Journal of Medical Entomology*, **45**, 1057–1063.

Simard F, Ayala D, Kamdem GC *et al.* (2009) Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology*, **9**, 17.

Slotman MA, Tripet F, Cornel AJ *et al.* (2006) Evidence for further subdivision within the M molecular form of *Anopheles gambiae*. *Molecular Ecology*, **16**, 639–649.

Tripet F, Toure YT, Taylor CE *et al.* (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology*, **10**, 1725–1732.

Turner TL, Hahn MW (2007) Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*. *Molecular Biology and Evolution*, **24**, 2132–2138.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *Plos Biology*, **3**, e285.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

White G (1974) *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **68**, 278–301.

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, **19**, 925–939.

Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

doi: 10.1111/j.1365-294X.2010.04532.x