Ecological Genomics of *Anopheles gambiae* Along a Latitudinal Cline: A Population-Resequencing Approach

Changde Cheng,* Bradley J. White,*,[1] Colince Kamdem,[†,‡] Keithanne Mockaitis,[§] Carlo Costantini,[‡] Matthew W. Hahn, and Nora J. Besansky*,[2]**

*Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, [†]Faculty of Sciences, University of Yaoundé I, Cameroon, [‡]Institut de Recherche pour le Développement (IRD), MIVEGEC (UMR UM1, CNRS 5290, IRD 224) and Laboratoire de Recherche sur le Paludisme, Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale (OCEAC), Yaoundé, Cameroon, [§]The Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405, and **Department of Biology & School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405

ABSTRACT The association between fitness-related phenotypic traits and an environmental gradient offers one of the best opportunities to study the interplay between natural selection and migration. In cases in which specific genetic variants also show such clinal patterns, it may be possible to uncover the mutations responsible for local adaptation. The malaria vector, *Anopheles gambiae*, is associated with a latitudinal cline in aridity in Cameroon; a large inversion on chromosome 2L of this mosquito shows large differences in frequency along this cline, with high frequencies of the inverted karyotype present in northern, more arid populations and an almost complete absence of the inverted arrangement in southern populations. Here we use a genome resequencing approach to investigate patterns of population divergence along the cline. By sequencing pools of individuals from both ends of the cline as well as in the center of the cline—where the inversion is present in intermediate frequency—we demonstrate almost complete panmixia across collinear parts of the genome and high levels of differentiation in inverted parts of the genome. Sequencing of separate pools of each inversion arrangement in the center of the cline reveals large amounts of gene flux (*i.e.*, gene conversion and double crossovers) even within inverted regions, especially away from the inversion breakpoints. The interplay between natural selection, migration, and gene flux allows us to identify several candidate genes responsible for the match between inversion frequency and environmental variables. These results, coupled with similar conclusions from studies of clinal variation in *Drosophila*, point to a number of important biological functions associated with local environmental adaptation.

UNCOVERING the genetic basis of adaptation to heterogeneous environments is a central goal of ecological genomics (Storz 2005). A direct approach to this problem entails quantitative trait locus mapping of experimental crosses to associate genetic variation with fitness-related traits. However, the direct approach relies on measurable phenotypic differences previously implicated in environmental adaptation. When experimental crosses are not feasible, or phenotypic trait differences and their fitness consequences are unknown or uncharacterized, an alternative indirect approach is required. In such cases, genome-wide scanning for regions of elevated sequence divergence between natural populations inhabiting different environments—but connected by gene flow—is a powerful strategy that can be used to search for gene–environment associations and thereby identify candidate loci potentially involved in the adaptive process (*e.g.*, Berry and Kreitman 1993; Bonin *et al.* 2006; Turner *et al.* 2008; Hohenlohe *et al.* 2010; Turner *et al.* 2010; Ellison *et al.* 2011; Kolaczkowski *et al.* 2011). This indirect population genomic strategy ("reverse ecology";

doi: 10.1534/genetics.111.137794
Manuscript received November 18, 2011; accepted for publication December 17, 2011
Supporting information is available online at http://www.genetics.org/content/suppl/2011/12/30/genetics.111.137794.DC1.
Sequence data from this article have been deposited with the ENA under accession no. ERP000955.
[1]Present address: Department of Entomology, University of California, Riverside, CA 92521.
[2]Corresponding author: Department of Biological Sciences, 317 Galvin Life Sciences Center, University of Notre Dame, Notre Dame, IN 46556-0369. E-mail: nbesansk@nd.edu

Y. Li *et al.* 2008) complements phenotype-based or candidate gene-based approaches, as it provides a comprehensive genome-wide view of divergence not otherwise possible.

Environmental gradients, such as those produced by shifts in altitude or latitude, impose spatially varying selective pressures on populations distributed along the gradient. Accompanying clinal variation of fitness-related traits and genotypes in these populations, as observed in a number of organisms, may reflect the action of natural selection (Endler 1977). One of the best-studied examples of clinal variation occurs in *Drosophila melanogaster* along the east coast of Australia, spanning tropical northern Queensland to temperate Tasmania (reviewed by Hoffmann and Weeks 2007). Fitness-related phenotypic traits such as body size, temperature and desiccation tolerance, and genetic polymorphisms—notably four cosmopolitan chromosomal inversions—vary clinally along the latitudinal gradient. Clinal patterns can arise from demographic processes, independent of selection. However, spatially varying selection is strongly implicated in maintaining the cline in Australia both because of the high rates of gene flow inferred from noncoding genetic markers (Kennington *et al.* 2003) and because parallel clinal patterns are found on different continents (De Jong and Bochdanovits 2003). This example and similar clines in other organisms provide a powerful context for implicating potentially adaptive genotypes and phenotypes.

As first revealed by Dobzhansky's pioneering work, the observation that chromosomal inversion frequencies are correlated with latitudinal (and seasonal) climatic transitions suggests that inversion polymorphisms in natural populations are maintained by intense selection pressure imposed by climate-related variables (Dobzhansky 1947; Krimbas and Powell 1992; Powell 1997; Schaeffer 2008). A number of models to explain the spread and maintenance of inversions hinge on the fact that they suppress recombination between the rearranged chromosomal regions and thus preserve linkage disequilibrium between favorable combinations of alleles (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). However, alternative gene arrangements are not completely isolated. Genetic exchange due to gene conversion and double crossovers, particularly away from the breakpoints toward the middle of an inversion, erodes linkage disequilibrium over time (Navarro *et al.* 1997; Andolfatto *et al.* 2001; Laayouni *et al.* 2003; Schaeffer and Anderson 2005). Accordingly, the ease of detecting footprints left by positive selection on loci inside of inversions depends upon the age of the inversion and the selection–recombination balance at loci conferring adaptation to local conditions.

Few studies have addressed the relative importance of chromosomal inversions *vs.* collinear genomic regions in facilitating adaptive evolution, and population genomic studies in a variety of organisms have sometimes yielded contradictory results (Hoffmann and Rieseberg 2008; Feder and Nosil 2009). The mosquito *Anopheles gambiae*, one of the major malaria vectors in Africa, is an excellent model system in which to address this question. The range of this species extends across most of tropical Africa, spanning various ecoclimatic zones from rainforest through savanna and sahel, although always in association with humans in rural or peri-urban settings. Polymorphic chromosomal inversions are abundant in *An. gambiae*, with seven common inversions segregating on chromosome 2 (Coluzzi *et al.* 1979; Coluzzi *et al.* 2002). These inversions are nonrandomly distributed across time and space in a manner that suggests their maintenance by selection. The frequencies of several inversions are positively correlated with aridity, such that frequencies peak and trough between dry and rainy seasons in a repeatable cyclical pattern (Coluzzi *et al.* 1979; Rishikesh *et al.* 1985; Petrarca *et al.* 1990; Toure *et al.* 1998). In addition, latitudinal clines of inversions in West and Central Africa run from mesic rainforest where the inversions are virtually absent, to xeric sahel where they are fixed, or nearly so (Coluzzi *et al.* 1979; Simard *et al.* 2009). Although the precise selective agents have not been established, thermal and desiccation stress are known threats to insects in arid environments (Gibbs 2002), and physiological tests on laboratory colonies of *An. gambiae* differing only in the arrangement of the 22-Mb *2La* inversion (*2La/a* or $2L+^{a}/+^{a}$) demonstrated that the inverted orientation confers greater resistance to thermal and desiccation stress, as predicted by its association with arid environments (Gray *et al.* 2009; Rocca *et al.* 2009). The complete genome sequence available for *An. gambiae* (Holt *et al.* 2002) facilitates population genomic strategies to identify loci responsible for environmental adaptations, whether inside or outside of chromosomal rearrangements.

In two previous studies, we employed gene-based microarrays in divergence mapping of single sympatric population samples of *An. gambiae* carrying alternative chromosomal arrangements on chromosome 2 (White *et al.* 2007a; White *et al.* 2009), with the goal of identifying candidate genomic targets of natural selection. These studies revealed regions of strikingly elevated divergence between alternative arrangements of inversion *2La* and provided compelling evidence that this inversion is maintained by selection, but failed to find significant divergence between most rearrangements on chromosome 2R. Because the microarray-based divergence mapping was geared toward identifying fixed or major frequency differences, and because the mapping platform was not a tiling array, existing differentiation between rearrangements on 2R may have escaped detection. Regions of elevated differentiation in collinear genomic regions was neither expected nor observed within sympatric population samples of the same *An. gambiae* molecular form (M or S).

Here, we expand the resolution and scope of our population genomics investigation of environmental adaptation in *An. gambiae* through whole genome resequencing of S form populations sampled near the ends and middle of a latitudinal gradient of aridity in Cameroon. Along this gradient, S form populations show steep clinal variation of

inversions *2La* and *2Rb*, but little genetic differentiation at 12 microsatellite markers on chromosome 3 (average F_{ST} = 0.0053) (Slotman *et al.* 2007), suggesting unrestricted gene flow. By examining spatial patterns of sequence variation, we identify candidate loci within and outside of chromosomal rearrangements that are potentially involved in local adaptation.

Materials and Methods

Population sampling and karyotyping

Indoor resting mosquitoes were sampled in September 2007 by insecticide spray sheet collection inside human dwellings, from seven villages spanning a latitudinal gradient from southern rainforest (3 52'N) to northern arid savanna (9 25'N) in Cameroon, Central Africa (Figure 1; for a more detailed ecogeographic description, see Simard *et al.* 2009). Specimens were preserved over desiccant in individual numbered microtubes. *An. gambiae* S form was identified by sequential morphological and molecular taxonomic methods (Gillies and De Meillon 1968; Santolamazza *et al.* 2004). Both molecular taxonomy and molecular karyotyping of the *2La* and *2Rb* inversions (White *et al.* 2007b; Lobo *et al.* 2010) were performed on DNA extracted from a single leg (Collins *et al.* 1987); the remaining carcass was held in reserve.

Population pools and genome resequencing

Four population pools were constructed for genomic resequencing, each consisting of DNA from 34 *An. gambiae* S form females (to avoid undersampling of the X chromosome in heterogametic males). Pooling was based on collection locality and karyotype of the *2La* inversion and was blind to karyotype of other inversions on 2R, including *2Rb*. Two equal pools—comprising homokaryotypes of the *2La* or $2L+^a$ arrangement, respectively—originated from the same central locality (Manchoutvi: 5 52'48" N, 11 06'49" E) with a high level of $2La/+^a$ polymorphism ($2L+^a$ frequency, 59%). The other two pools originated from localities at opposite ends of the cline, where $2La/+^a$ polymorphism is low. At the southern end where $2L+^a$ predominates, a pool of $2L+^a/+^a$ homokaryotypes was constructed from Nkometou III (3 51'56" N, 11 30'56" E; $2L+^a$ frequency, 88%). At the northern end where *2La* predominates, a single pool of *2La/a* homokaryotypes was constructed from two nearby villages (Wack and Bini: respectively, 7 41'02" N, 13 32'56" E and 7 23'55" N, 13 33'02" E; average $2L+^a$ frequency, 6%) (Figure 1).

DNA was isolated from individual carcasses using the Wizard SV 96 genomic DNA purification system (Promega, Madison, WI). To prepare pools, 50 ng of DNA from each of 34 female mosquitoes (68 chromosomes) was combined (1.7 μg total). Sequencing libraries were constructed from each of the four pools by the Center of Genomics and Bioinformatics (CGB) at Indiana University, Bloomington, according to standard Illumina protocols. Each library was sequenced on an Illumina Genome Analyzer IIx at the CGB to generate 72-bp single-end reads, with the goal of achieving ~10× average genome coverage (over 260 Mb) while avoiding oversampling of the same chromosome sequence in the pools of 68 chromosomes (Futschik and Schlotterer 2010). Across all four libraries, this produced 245,586,259 reads (approximately 18.7 Gb of sequence data). These reads are available from the Short Read Archive (SRA) of the European Nucleotide Archive (ENA) under accession no. ERP000955.

Read alignment

Illumina sequencing reads were aligned to the AgamP3 (PEST) assembly of the *An. gambiae* genome (http://www.vectorbase.org; Lawson *et al.* 2009). With the goal of maximizing the number of mapped reads, we conducted two separate alignments to AgamP3 using different programs, and compared metrics between the two. The first program, Mapping and Assembly with Qualities (MAQ, v. 0.7.1; H. Li *et al.* 2008), is computationally efficient, but was developed for genomes with relatively low levels of polymorphism (*e.g.*, human) and it does not allow for mapping of reads with insertions or deletions (indels). Among the 120,297,539 reads uniquely mapped by MAQ (reads with two or more equally likely positions were excluded), the average mismatch rate to the AgamP3 reference was 2.8%. Average genome coverage was 39.7-fold summed across all four pools.

The second program, SHort Read Mapping Package (SHRiMP, v. 1.3.0; Rumble *et al.* 2009), can capture higher levels of polymorphism including indels, but is computationally intensive. To reduce runtime, we parallelized SHRiMP jobs through Condor, a distributed batch computing system at the University of Notre Dame. Reads mapped with SHRiMP were processed with the PROBCALC utility. After excluding reads with multiple equally good matches to the reference (alignments with pchance <0.05, or normodds <0.8), 137,185,394 uniquely mapped reads were piled up to the reference genome using custom Perl scripts. The average mismatch and indel rates between reads and reference were 2.6 and 0.4%, respectively. The 44.9-fold average genome coverage achieved by SHRiMP across the four pools was ~10% higher than MAQ. Supporting information, Figure S1 (left), shows read coverage for MAQ and SHRiMP across the four pools by chromosome arm. In Figure S1, right, a detailed plot of read coverage along the length of chromosome 2L based on the northern (*2La/a*) population pool indicates the substantial increase in coverage achieved by SHRiMP in the rearranged region of 2L, where a much higher level of read mismatch (to the uninverted $2L+^a$ AgamP3 reference) is expected. Average read coverage inside the rearranged region is 12× with SHRiMP compared to 8.5× with MAQ for the northern pool. Because SHRiMP mapping provided higher read coverage both inside and outside inversions, all downstream analysis of genetic diversity and differentiation was based on SHRiMP alignments.

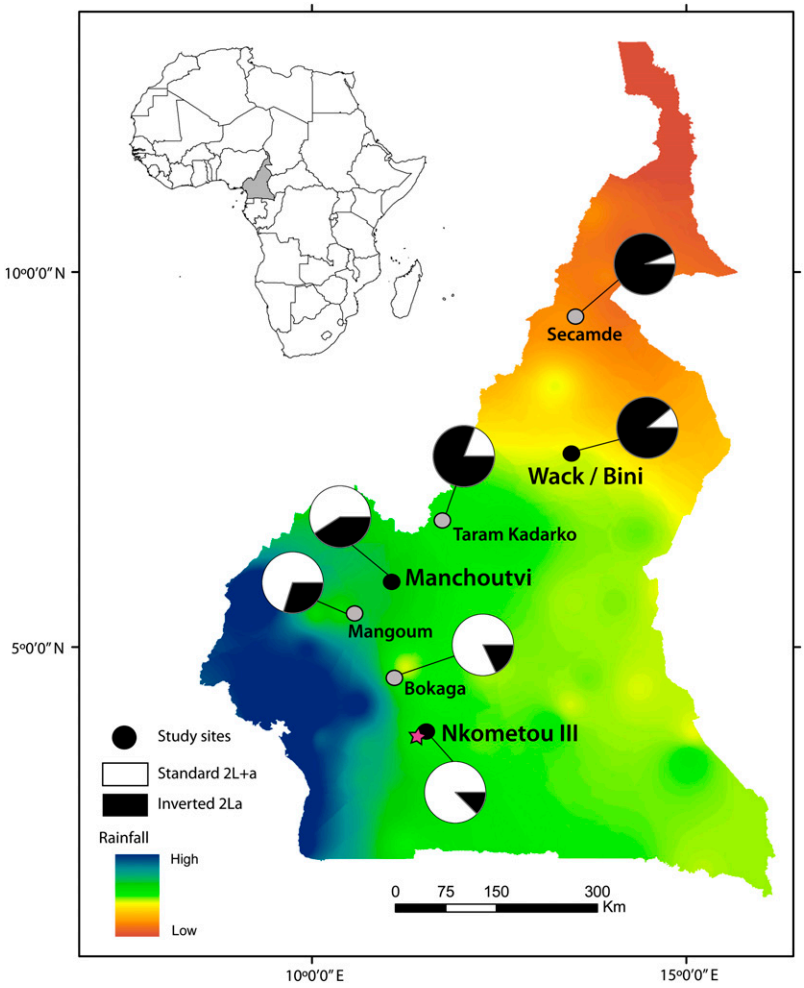

Figure 1 Sampling sites along a latitudinal transect in Cameroon. Mean annual precipitation, from high (blue) to low (orange), is based on data from FAO SDdimensions (http://www.fao.org/sd/2002/EN1203a_en.htm). Pies show the frequency of the standard (*2L+*^{a}, white) and inverted (*2La*, black) arrangement of inversion *2La* in *An. gambiae* S populations sampled at each locality (circles). Black circles indicate the three localities chosen for population resequencing. The capital Yaoundé is represented by a star.

Data filtering

Only uniquely aligned reads were retained for analysis. To be considered, a site was required to have at least 10× read coverage in each population being compared, but no more than 30× coverage (to mitigate against repetitive sequences). For window-based analyses, only windows containing a minimum number of sites meeting coverage thresholds were retained. The minimum number of sites required per 1-kb window (317) was arbitrarily set on the basis of empirical distribution, to exclude windows containing the smallest 25% of usable sites. Both site-based and window-based analyses excluded low recombination regions of the genome designated as heterochromatic (Sharakhova *et al.* 2010), including pericentromeric heterochromatin (X—20,009,764-24,393,108; 2L—1-2,431,617; 2R—58,984,778-61,545,105; 3L—1-1,815,119; 3R—52,161,877-53,200,684) and intercalary heterochromatin (2L—5,078,962-5,788,875; 3L—4,264,713-5,031,692; 3R—38,988,757-41,860,198).

SNP identification and frequency inference

To reduce the effect of random sequencing error, singleton polymorphisms (those found only in a single read in a single population) were discarded. Although this treatment should also remove true low-frequency polymorphisms, it should have little consequence on the ability to detect the most extreme frequency differences between populations—the focus of this study. For polymorphisms supported by at least two reads, SNP frequency estimates were weighted according to their quality scores (given by the Illumina GA pipeline software), using Holt *et al.*'s (2009) Equation 6.

Population genomic and outlier analysis

Following Akey *et al.* (2010) and Kolaczkowski *et al.* (2011), we adopted an empirical outlier approach to identify exceptionally diverged regions of the genome that are potentially affected by selection. In comparing populations sampled at opposite ends of the cline, we found that average levels of divergence between rearranged (*2La*, *2Rb*) genomic regions were drastically higher than those between collinear genomic regions (see Figure 2). Given our interest in identifying outliers outside as well as within the *2La* and *2Rb* rearrangements, outlier analysis was performed separately for each rearrangement and for the remaining (collinear) regions exclusive of heterochromatin. For each of these three data partitions, measures of diversity and divergence were estimated from nonoverlapping 1-kb windows. Specifically, bias-corrected estimates of Tajima's π for each 1-kb window were calculated

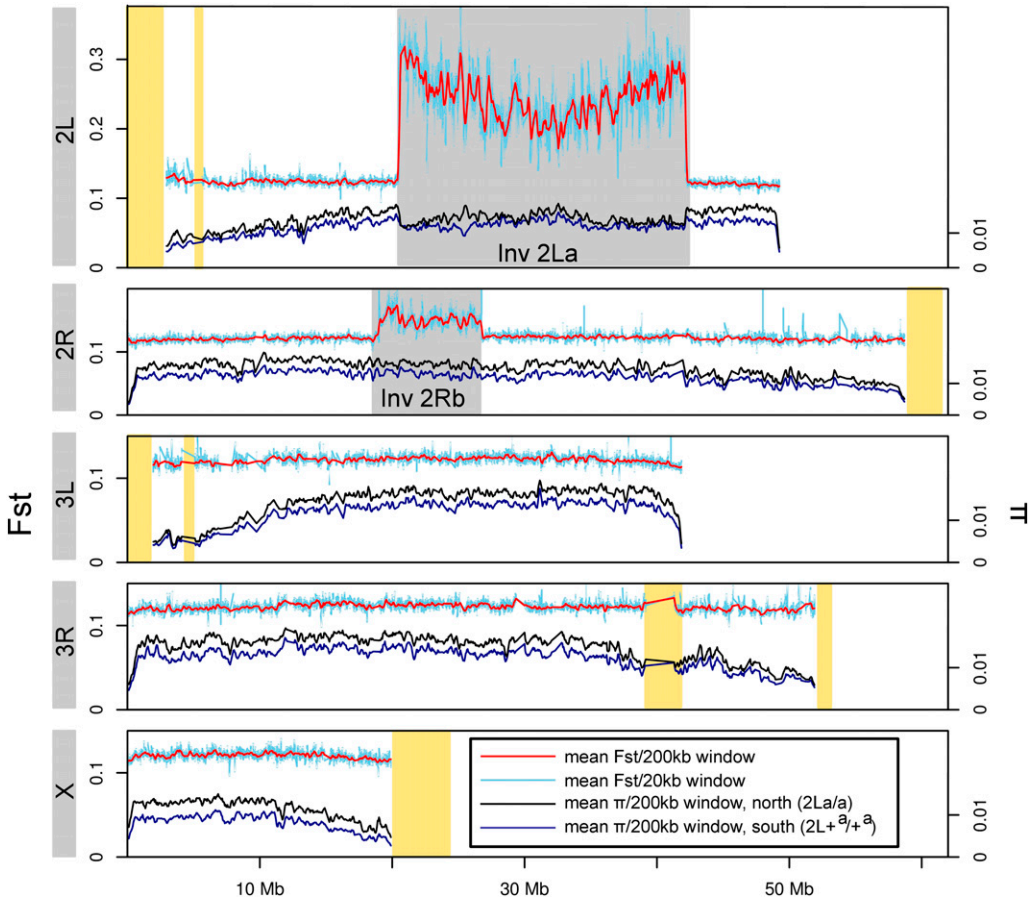

Figure 2 Divergence (F_{ST}) between and diversity (π) within *An. gambiae* S populations from opposite ends of the latitudinal cline in Cameroon. Across each chromosome arm, mean divergence is plotted over 200-kb windows slid every 50 kb (red), and 20-kb windows slid every 5 kb (light blue). Mean diversity is plotted for 200-kb windows slid every 50 kb for Nkoumetu in the south (*2La*/a; dark blue) and Wack/Bini in the north ($2L+^a/+^a$; black). Gray and yellow shaded boxes indicate chromosomal rearrangements or heterochromatic regions, respectively. Only windows containing ≥317 sites meeting coverage thresholds are plotted, causing reduced representation of heterochromatic regions.

following Futschik and Schlötterer (2010). These estimates were used to derive F_{ST} values, calculated as $F_{ST} = (\pi_{Between} - \pi_{Within})/\pi_{Between}$ (Hudson *et al.* 1992) and averaged over all segregating SNPs in each 1-kb window. For each data partition, outlier windows were defined as those with mean F_{ST} values falling in the top 1% of the empirical distribution. For some analyses, individual outlier SNPs were defined as those with F_{ST} values falling in the top 0.1% of the empirical distribution (collinear region and *2Rb* rearrangement) or $F_{ST} = 1$ (*2La*).

Copy-number variation

To identify possible copy-number differences between the two populations at opposite ends of the cline, average read depth coverage was estimated in 1-kb windows across nonheterochromatic regions of the genome. Window-based ratios of read depth between the two populations were normalized by the median genome-wide ratio and plotted following arctan transformation. We considered windows in the top 0.5% or bottom 0.5% of the empirical distribution of normalized depth ratios to represent candidate copy-number differences (Kolaczkowski *et al.* 2011).

Functional clustering

Lists of candidate genes were assembled from the set that overlapped the outlier F_{ST} or read depth ratio 1-kb windows. For this purpose, a gene was defined as the predicted transcribed region plus 1 kb upstream and downstream. The resulting gene lists were explored for possible functional relationships based on annotation profiles built from terms derived from multiple sources [*e.g.*, Gene Ontology (GO), SMART and InterPro Domains, SwissProt/Uniprot, and PIR keywords], using the DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/) (Huang *et al.* 2009). We employed DAVID's Functional Annotation Clustering utility with default settings to identify groups of genes ("annotation clusters") whose annotation profiles suggest common function. The enrichment score assigned to each annotation cluster represents the relative importance of that functional gene group on the basis of the fraction of its members associated with highly enriched annotation terms. It is measured by the geometric mean of the EASE Scores (a modified Fisher exact *P*-value) associated with each enriched annotation term in the gene group (Hosack *et al.* 2003; Huang *et al.* 2007) and is intended to rank the importance of the groups, rather than to provide a rejection/acceptance threshold customary of a traditional statistical analysis. For this reason, enrichment scores are presented in the form of minus log-transformed geometric means instead of an absolute *P*-value (Huang *et al.* 2007).

Results

We resequenced whole genomic DNA from four population pools each composed of 34 *An. gambiae* S form mosquitoes

of known 2L homokaryotype (*2La/a* or *2L+^{a}/+^{a}*), derived from three localities along an arid-to-mesic latitudinal gradient in Cameroon: two from the extremes of the gradient and one in the center (Figure 1). Because the climatic gradient is associated with a steep cline in *2La* inversion frequency, from near fixation in the arid north to near absence in the mesic south, our study design included samples of each homokaryotype from either endpoint and in the center. Specifically, we sampled a homokaryotypic *2La* population in the north and a homokaryotypic *2L+^{a}* population in the south and separately sequenced pools of homokaryotypic *2La* and *2L+^{a}* individuals from the central population (where heterokaryotypes are common in nature).

Two flow cells of Illumina GA IIx resequencing produced ~45-fold coverage of the genome across all four population pools (12.2× and 9.6× from the north *2La/a* and south *2L+^{a}/+^{a}* populations; 10.3× and 12.8× from the central *2La/a* and *2L+^{a}/+^{a}* populations). After retaining only uniquely aligned reads, filtering based on read coverage, and eliminating 1-kb windows with insufficient data meeting coverage thresholds (*Materials and Methods*), the average 1-kb window contained 779 sites. In the combined populations, 17,077,264 nonsingleton SNPs were detected outside of regions designated as centromeric or intercalated heterochromatin. Importantly, population-pairwise estimates of F_{ST} based on 1-kb windows indicate only moderate genome-wide differentiation between any population pair (Table S1), and differentiation between the northern and southern samples (F_{ST} = 0.123 excluding rearranged regions) is only marginally greater than that between subsamples of a single central population (F_{ST} = 0.114). Taken with the absence or very low level of significant microsatellite differentiation between comparable population samples of *An. gambiae* S from Cameroon (Slotman *et al.* 2007), these data suggest sufficient genetic connectivity to merit a reverse ecology approach for mapping candidate genes underlying local adaptation.

Genomic patterns of divergence between endpoint populations

Nucleotide divergence between populations sampled from the northern and southern ends of the climatic gradient and nucleotide diversity within these populations were estimated from nonoverlapping 1-kb windows across the euchromatic chromosome arms (Figure 2; Table S2). Excluding the two rearranged regions on chromosome 2, mean F_{ST} is relatively low (0.123) and similar among all five chromosome arms. In contrast, mean F_{ST} is strikingly elevated—although to different extents—inside the regions spanned by the chromosome 2 rearrangements (F_{ST} inside *2La* = 0.247; F_{ST} inside *2Rb* = 0.149). Elevated F_{ST} does not persist outside of the rearrangements beyond 20 kb from the breakpoints, in contrast to the larger extent of divergence (1–2 Mb) observed outside the breakpoints of inversions distinguishing *Drosophila pseudoobscura* and *D. persimilis* (Machado *et al.* 2007).

More pronounced differentiation between alternative arrangements of *2La* relative to alternative arrangements of *2Rb* has been noted previously in natural populations of *An. gambiae* (White *et al.* 2009). However, in the present data set the distinction between levels of divergence observed in the two rearrangements is also partly due to the fact that population pooling took only the *2La* karyotype into account and was blind to *2Rb* karyotype. Molecular karyotyping of *2Rb* (Lobo *et al.* 2010) performed on individual mosquito DNA aliquots after Illumina libraries had been prepared from pooled DNA revealed a frequency of *2Rb* ranging from ~7% in the southern to ~90% in the northern population pools, respectively (and ~67% in the central locality). Thus, as expected (Coluzzi *et al.* 1979; Lee *et al.* 2009), *2Rb* is clinally distributed along the arid-mesic latitudinal gradient in a pattern that parallels the *2La* inversion. These data also indicate that although alternative *2Rb* arrangements predominate at opposite ends of the cline, both arrangements are present in the two population pools, presumably resulting in some dampening of observed levels of divergence in the region spanned by *2Rb* between northern and southern population samples.

Across the genome, nucleotide diversity is consistently lower in the southern than in the northern population (Figure 2 and Table S2). However, the general trends are very similar between the two. Most notable are markedly lower diversity toward telomeric and especially centromeric ends of the chromosomes, and lower diversity inside the *2La*, but not the *2Rb*, rearrangement relative to flanking chromosomal regions.

Given the heterogeneous levels of divergence in collinear and rearranged regions of 2L and 2R, and our interest in identifying elevated divergence within each of these three regions, we treated all three as separate data partitions when assessing outlier 1-kb windows of F_{ST}. Histograms of F_{ST} values for windows in each data partition are provided in Figure 3. The threshold F_{ST} values for windows in the top 1% of the three distributions differed greatly, and the rank order was: collinear genome (0.163) < *2Rb* (0.234) < *2La* (0.431).

Our ability to identify candidate genes for local adaptation depends upon the degree of linkage disequilibrium between the polymorphisms most strongly associated with populations at the endpoints of the climatic gradient (*i.e.*, SNPs with the highest F_{ST} values between our northern and southern population pools). As haplotype information is lost in pooled sequencing, we approached this question by assessing the rate of decay of F_{ST} with increasing distance from a focal SNP (Turner *et al.* 2010), treating each of the three data partitions separately. Focal SNPs for this analysis were defined as those carrying the highest 0.1% of F_{ST} values for the collinear genome and the *2Rb* rearrangement (respectively, 9458 SNPs with $F_{ST} > 0.43$ and 560 SNPs with $F_{ST} > 0.77$), and those with $F_{ST} = 1$ for the *2La* rearrangement (25,915 SNPs). Beginning with a window size of 1 bp (the focal SNP itself), windows centered on each focal SNP

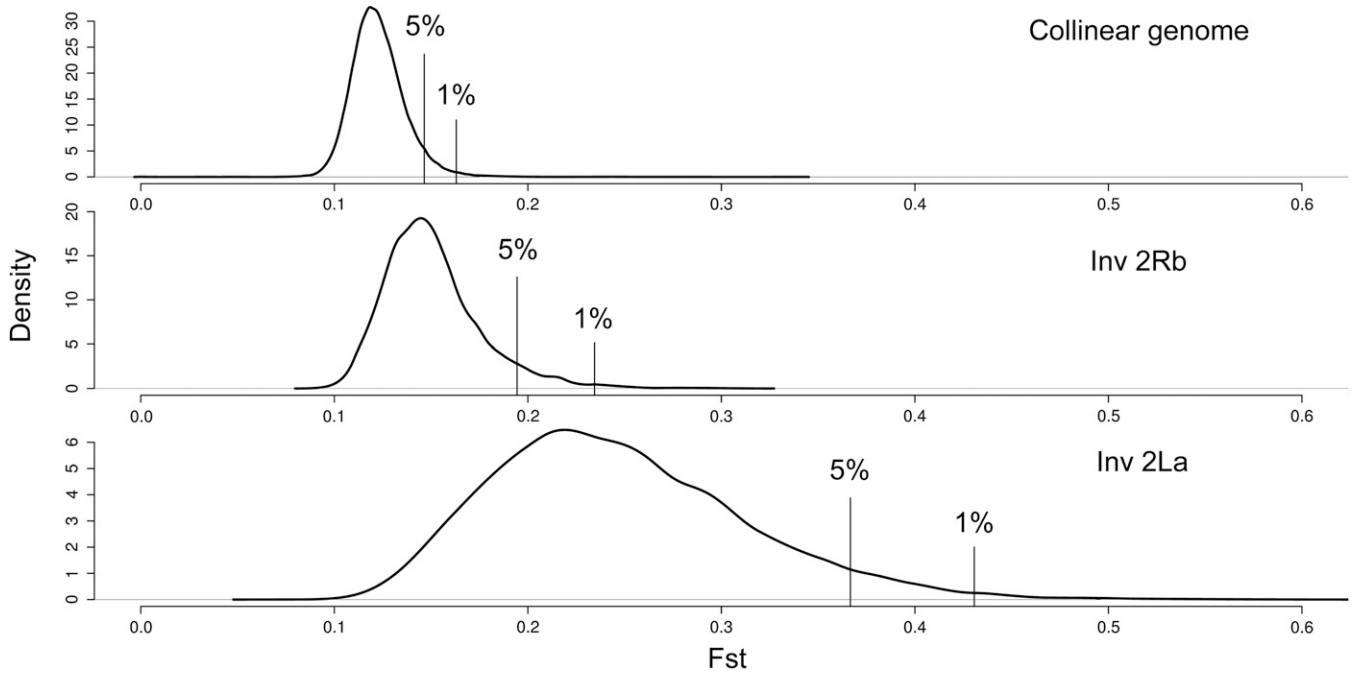

Figure 3 Distribution of F_{ST} values for 1-kb windows spanning the collinear euchromatic genome, the *2Rb* and the *2La* rearrangements (between north and south populations). Positions in the distributions marking the top 5% and top 1% F_{ST} values are indicated by vertical lines.

were incrementally increased in size up to 20,000 bp. Mean F_{ST} was recalculated per window and averaged across windows of the same size. As seen in Figure 4, F_{ST} declines precipitously within 500 bp, not only in the collinear genome but also inside both rearrangements, suggesting that the most differentiated SNPs are not colocalized.

Clinal patterns of divergence

We focused on the three sets of SNPs whose north–south F_{ST} values were in the top 0.1% for the *2Rb* and collinear regions, and maximal ($F_{ST} = 1$) for *2La* (the same SNPs used to explore the decay of F_{ST} with increasing distance, described above). Allele frequency at each SNP, measured with respect to the major allele in the southern population, was estimated for northern, southern, and central populations. In the case of the central population, allele frequencies in each genomic region were estimated for both (*2La*/a and $2L+^{a}/+^{a}$) subsamples, as these were sequenced separately. As shown in Figure 5, the SNP frequencies in all three genomic regions—*2La*, *2Rb* and collinear—are clinal; mean frequencies in the central population are intermediate to the end populations.

A cross-comparison of patterns among rearranged and collinear regions demonstrates the power of following the separate fates of the *2La*/*a* and $2L+^{a}/+^{a}$ samples in the center of the cline. Although differences in allele frequencies between the two samples are significantly different by Wilcoxon signed rank test in each case (P-value $< 2.2 \times 10^{-16}$), the magnitude of difference is substantially less for collinear as compared to rearranged regions. SNPs in the collinear genome that differentiate these mosquitoes at opposite ends of the cline are effectively homogenized between alternative karyotype subsamples in the central population (Figure 5A). These data indicate that the two groups of mosquitoes carrying alternative homozygous arrangements of *2La* are not genetically isolated from each other, and in fact exchange alleles across the majority of the genome. By contrast, and in

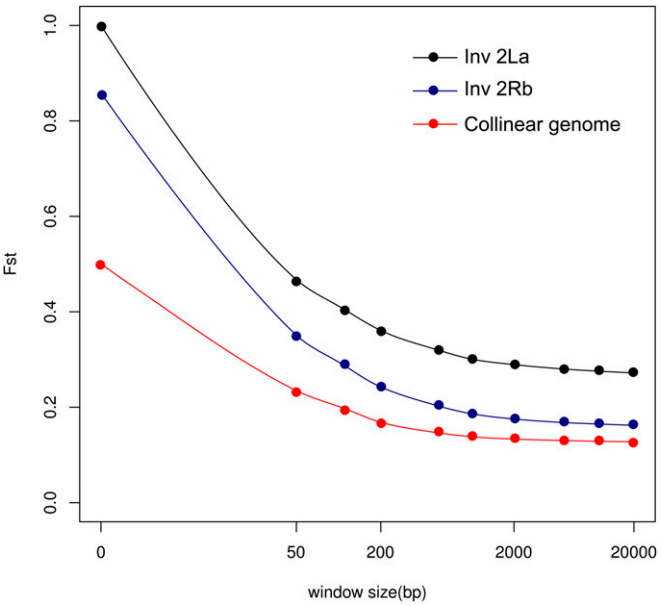

Figure 4 The decay of average F_{ST} with increasing window size for windows centered on the most differentiated SNPs between northern and southern Cameroon populations in the collinear genome (red), the *2Rb* rearrangement (blue), and the *2La* rearrangement (black). Note that the *x*-axis is on a logarithmic scale.

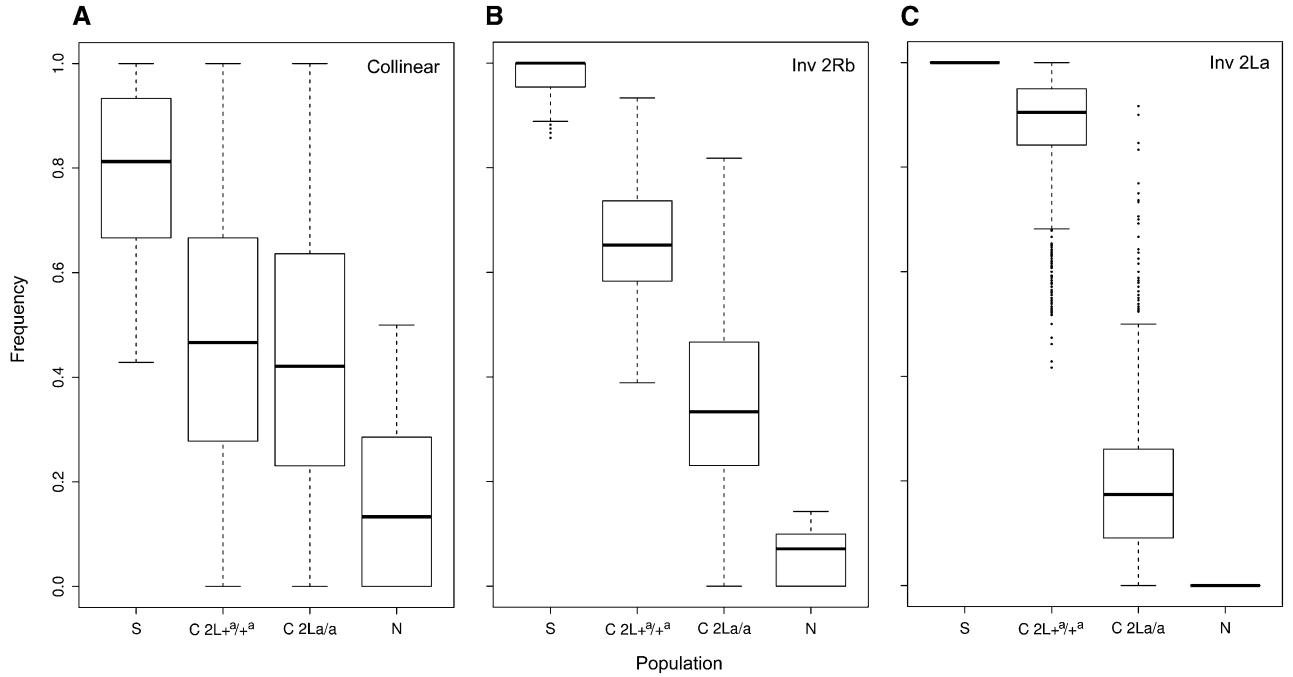

Figure 5 Individual SNP frequencies along the latitudinal gradient in Cameroon, based on loci that most strongly differentiated populations at opposite ends of the gradient. Frequencies are plotted with respect to the allele predominating in the southern population in (A) the collinear genome, (B) the *2Rb* rearrangement, and (C) the *2La* –75th percentile); the horizontal line marks the median. Upper (lower) whiskers include the maximum (minimum) values unless the distance from the first (third) quartile exceeds 1.5× that of the IQR. Outlier values smaller or larger than the whiskers are indicated by dots. N, northern population; S, southern population; C, central population, partitioned by karyotype into *2La/a* and $2L+^{a}/+^{a}$ homozygotes.

keeping with the association between *2Rb* and *2La* karyotypes along the cline in Cameroon, *2Rb*-associated SNPs appear more likely to be found in *2La/a* samples in the central population (while $2R+^{b}$-associated SNPs are more likely in $2L+^{a}/+^{a}$ samples), despite independent segregation of the two inversions (Figure 5B).

The most compelling result concerns the ~26,000 SNPs inside the *2La* rearrangement that are completely fixed for alternative states at opposite ends of the cline. In the center of the cline, considering *2La* and $2L+^{a}$ chromosomes separately, there is a clear shift in allele frequencies on both arrangements such that many SNPs are no longer fixed between alternative karyotypes and some show radical changes in frequency (Figure 5C). Gene flux (the transfer of alleles between alternative arrangements by crossing over and gene conversion) is the most reasonable explanation for this pattern, given that it occurs orders of magnitude more frequently than new mutations (Navarro *et al.* 1997; Andolfatto *et al.* 2001; Laayouni *et al.* 2003). Of particular note is the relatively large number of *2La* and $2L+^{a}$ SNPs whose frequencies are extreme (indicated by dots in 5C) relative to the distribution of frequency values for the central population; these represent outlier SNPs that have been homogenized between inversion arrangements. In Figure 6, the location and density of these outlier SNPs in the center of the cline can be compared to the distribution and density of all fixed SNP differences between alternative *2La* arrangements at opposite ends of the cline. The outlier SNPs are distributed across the length of the inversion, but not uniformly. They are more rare toward the breakpoints and more frequent in the center of the inversion. This pattern is consistent with gene flux, given that gene conversion is expected to predominate closer to the breakpoints while double crossovers and gene conversion both operate in the center of an inversion (Navarro *et al.* 1997; Andolfatto *et al.* 2001; Laayouni *et al.* 2003). If gene flux explains the admixture of *2La*- and $2L+^{a}$-associated alleles in the central population, as seems likely, the implication is that selection is maintaining the clinal pattern of differentiation at many SNPs against the potent homogenizing forces of gene flow and gene flux.

Genic patterns of divergence

To obtain a genic view of elevated differentiation between northern and southern populations, we compiled three sets of genes that overlapped 1-kb windows falling within the top 1% of F_{ST} values in the rearranged (*2La*, *2Rb*) and collinear regions of the genome (listed in Table S3, Table S4, and Table S5 with their putative *Drosophila* orthologs obtained from the Ensembl Metazoa Genes 10 database). Considering the gene sets separately, we subjected each to a functional annotation clustering analysis using the DAVID software package (Huang *et al.* 2009), with the goal of condensing the gene lists into groups potentially associated with common biological processes or functions.

The *2La* rearrangement spans 1281 genes, of which 52 overlapped the most divergent 1-kb windows. Functional

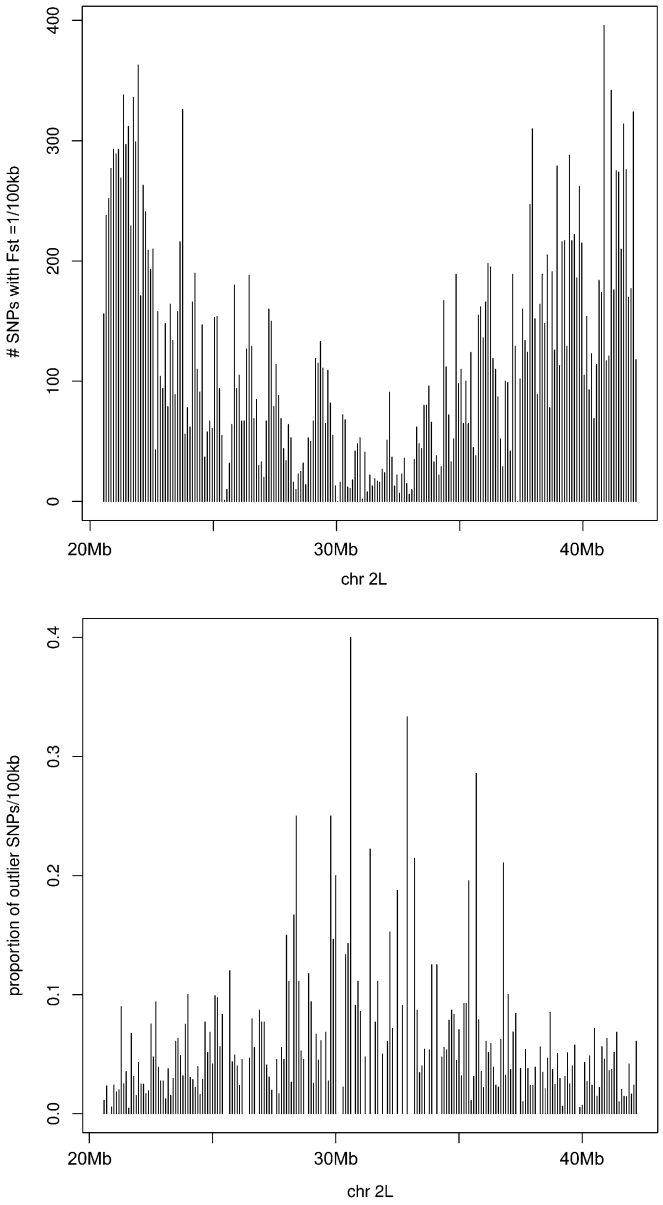

Figure 6 The position and density of SNPs inside the *2La* rearrangement with fixed differences between *2La* and *2L+^{a}* at opposite ends of the cline. Counts are based on nonoverlapping 100-kb windows slid between the breakpoints of *2La*. Top: The total set of SNPs with F_{ST} = 1 between northern and southern populations. Bottom: The outlier SNPs in the central *2La/a* and *2L+^{a}/+^{a}* populations (see Figure 5C).

annotation clustering attempted on the 52 genes revealed four small groups whose enriched annotation terms classified them as cuticle proteins (annotation cluster 1), serine/threonine protein kinases (2), and ion channels/GPCRs (clusters 3 and 4 contained the same 8 genes) (Table 1). Annotation cluster 1 contains mainly cuticle protein (CPR) genes of the RR-2 family, which are often found in close proximity in different regions of the *An. gambiae* genome. Indeed, the physical cluster of CPR genes inside *2La* is the largest, containing ~40 CPR genes, which overlap a ~1-Mb region previously identified through microarray-based divergence mapping as one of the most diverged between alternative arrangements of *2La* (White *et al.* 2007a). Among the serine/threoine protein kinase genes in cluster 2 with identifiable homologs are several putative *Drosophila* orthologs that regulate aspects of growth, development, and innate behavioral responses. These include *cyclin-dependent kinase 4* and *nimA-like kinase* (growth regulation and mitotic control), *nemo* (specification of polarity and development of wing size/shape), *frayed* (axon ensheathment), and *alan shepard* (geotaxis). Another is *JIL-1*, a gene whose protein kinase product is proposed to reinforce the status of active chromatin through phosphorylation of histone H3 at serine 10 (Regnard *et al.* 2011). Divergence between northern and southern populations mapped onto the *An. gambiae* gene model of *JIL-1* (AGAP006094) is illustrated in Figure 7A. Within clusters 3 and 4 are ion channel-related genes of the potassium voltage-gated channel family (*KCNQ* and *shal2* orthologs) implicated in nervous system development and function. In addition, there are chemosensory receptors: three originally classified as G protein-coupled receptors (GPCRs) in the gustatory receptor family (GPRGR29, GPRGR30, GPRGR31; Figure 7B) and another ionotropic glutamate receptor (GPRMGL1). Unsurprisingly, the DAVID tool did not cluster all 52 genes overlapping the top 1% F_{ST} windows. Manual appraisal of the complete *2La* gene list in conjunction with putative *Drosophila* ortholog assignments uncovered several unclustered candidates with potential roles in the regulation of nervous system development and function, roles that cut across some of the DAVID annotation clusters (*diacylglycerol kinase 1*, *follistatin*, *multiplexin*, *pebble*, *peptidylamidoglycolate lyase 2*, *still life*, *syntrophin-like 2*, and *unc-13-4A*). Also unclustered, but functionally related to *JIL-1* in the context of chromatin regulation of transcription, was *asf1* (a histone chaperone with a role in SWI/SNF-mediated chromatin assembly and remodeling).

The smaller *2Rb* rearrangement spans only 548 genes. Among the 24 genes overlapping the most diverged 1-kb windows, a single annotation cluster (5) of 4 genes shared annotation terms suggesting that they could be members of the immunoglobulin superfamily. Members of this superfamily play important roles in development, and cell–cell adhesion and communication (Vogel *et al.* 2003). Only two genes in cluster 5 had named homologs in *Drosophila* (*klingon*, *defective proboscis response 18*), and only *klingon* has a known functional role in neurogenesis, participating in the development of the R7 receptor neuron. Two other candidate genes that were not clustered also potentially participate in neural development on the basis of their *Drosophila* orthologs: the homeobox gene *rough* (required in photoreceptors R2 and R5 for inductive interactions in the developing eye) and *twisted*. Figure 7C shows divergence in the vicinity of AGAP002628, a gene of unknown function in cluster 5.

Of the considerably larger set of genes (11,262) in the collinear (euchromatic) genome, 485 overlapped the top 1% F_{ST} windows and were placed into a number of different

Table 1 Functional annotation clusters of genes overlapping the top 1% windows of F_{ST} between northern and southern populations in rearranged (*2La*, *2Rb*) and collinear genome regions

Genome partition	Annotation cluster	Representative annotation term(s)	Gene count	Enrichment score
2La	1	Insect cuticle protein	5	1.56
	2	Serine/threonine protein kinase	9	1.52
	3	Ion channel/transport	8	1.51
	4	GPCR	8	1.17
2Rb	5	Immunoglobulin-like fold	4	2.50
Collinear	6	Cell-surface receptor-linked signal transduction	63	2.97
	7	Calcium-dependent membrane targeting	6	1.84
	8	Transcription regulator/homeodomain	32	1.45
	9	GPCR/neurotransmitter receptor activity	7	1.30
	10	ETS domain/winged helix repressor DNA binding	3	1.12
	11	Postembryonic development/morphogenesis	14	1.03

clusters. Six of the most important annotation clusters are given in Table 1. Interestingly, these clusters and their component genes reflect many of the same functional themes evoked by genes inside the rearrangements: development (including neural development and differentiation in the eye and elsewhere), receptor-mediated signaling, and transcriptional regulation. Clusters 6, 7, and 9 contain genes whose products putatively function as ion channels and receptors, analogous to clusters 3–4 for the *2La* rearrangement: genes encoding odorant and gustatory receptors, an insulin-like receptor, a muscarinic acetyl choline receptor, and two transient receptor potential cation channels (*trpl* and *TrpA1* orthologs). Scattered across annotation clusters, or in some cases not clustered, are genes whose *Drosophila* orthologs are involved in the signaling pathways Wnt (*e.g.*, *Wnt 2*; *frizzled 4*), EGF (*star*), and TGF-B (*Smad on X*), or genes whose products control response to ecdysone (dopamine/ecdysteroid receptor *DopEcR*, ecdysone receptor-interacting factor *smrter*, ecdysone-induced protein 74E *eip74E*). Of particular note is another chromatin-modifying gene encoding a histone deacetylase (*HDAC4*).

One intergenic region in the collinear genome captured our attention, because it contains the sole instance of a SNP with $F_{ST} = 1$ between northern and southern populations outside of the two rearranged regions (Figure 7D). Located on chromosome 3R, this SNP is surrounded by SNPs with far lower F_{ST} values (not exceeding $F_{ST} = 0.4$) and is in the neighborhood of two genes. It is ~2 kb upstream of a tRNA (Arg) and ~4 kb upstream of a zinc finger protein whose predicted *Drosophila* ortholog *earmuff* (*erm*) is involved in the regulation of neurogenesis—a functional theme enriched among the set of genes that overlap the most diverged windows genome-wide. However, it is currently unknown whether this SNP is located in a regulatory region.

Aside from inferences based on functional annotation clustering, we mined a newly available gene expression atlas of sex and tissue specificity in *An. gambiae* (Baker *et al.* 2011) for additional clues about the nature of candidate genes. Overall, the 620 candidates were interrogated by 761 probes in the MozAtlas microarray experiments (http://www.tissue-atlas.org). However, gene expression was not detected for almost half of these probes (>57% of those inside rearrangements were not detected as expressed). In Table S3, Table S4, and Table S5, we have indicated which candidate genes had at least one probe whose expression was detected, and of those, which had gender-biased expression or expression limited to one sex and one tissue. Although it is difficult to make robust quantitative statements on the basis of these data, we note a similar level of gender-biased expression as that reported by MozAtlas for the genome as a whole (~43% of expressed probes), and an intriguing enrichment of sex- and tissue-specific expression (29% of expressed probes), particularly testis-specific expression.

Copy number variation

In addition to nucleotide divergence, we also scanned the genome for potential genic regions of copy-number variation (CNV). Two caveats should be noted about detecting CNVs in our data set. Our data-filtering excluded reads with multiple equally good alignments, a precaution intended to minimize problems in distinguishing orthology from paralogy in repetitive DNA, but also a step that may have removed loci with CNVs. Additionally, if sequence divergence from the reference genome is more extreme for one of the endpoint populations, the overall effect on read mapping might mimic the pattern expected for CNVs (*i.e.*, lower coverage in one population relative to the other), leading to false positives. Bearing in mind these caveats, we identified outlier 1-kb windows in the distribution of normalized read depth coverage (0.5% upper and lower tails) (Figure S2). The set of 415 genes overlapping windows of unusually skewed read coverage were compiled (Table S6) and submitted for functional annotation clustering using DAVID (Table S7). Although the resulting candidate CNV genes do not generally overlap the genes associated with high F_{ST} 1-kb windows, similar functional themes emerged from the most important clusters, including transmembrane receptor activity/tyrosine protein kinase signaling, development/morphogenesis, and EGF-responsive genes. In Figure 7E is a representative example from the putative *Methuselah receptor 4* gene inside the *2La* rearrangement, showing SNP-based divergence and normalized fold coverage differences in both populations. Surprisingly, there is some suggestion of

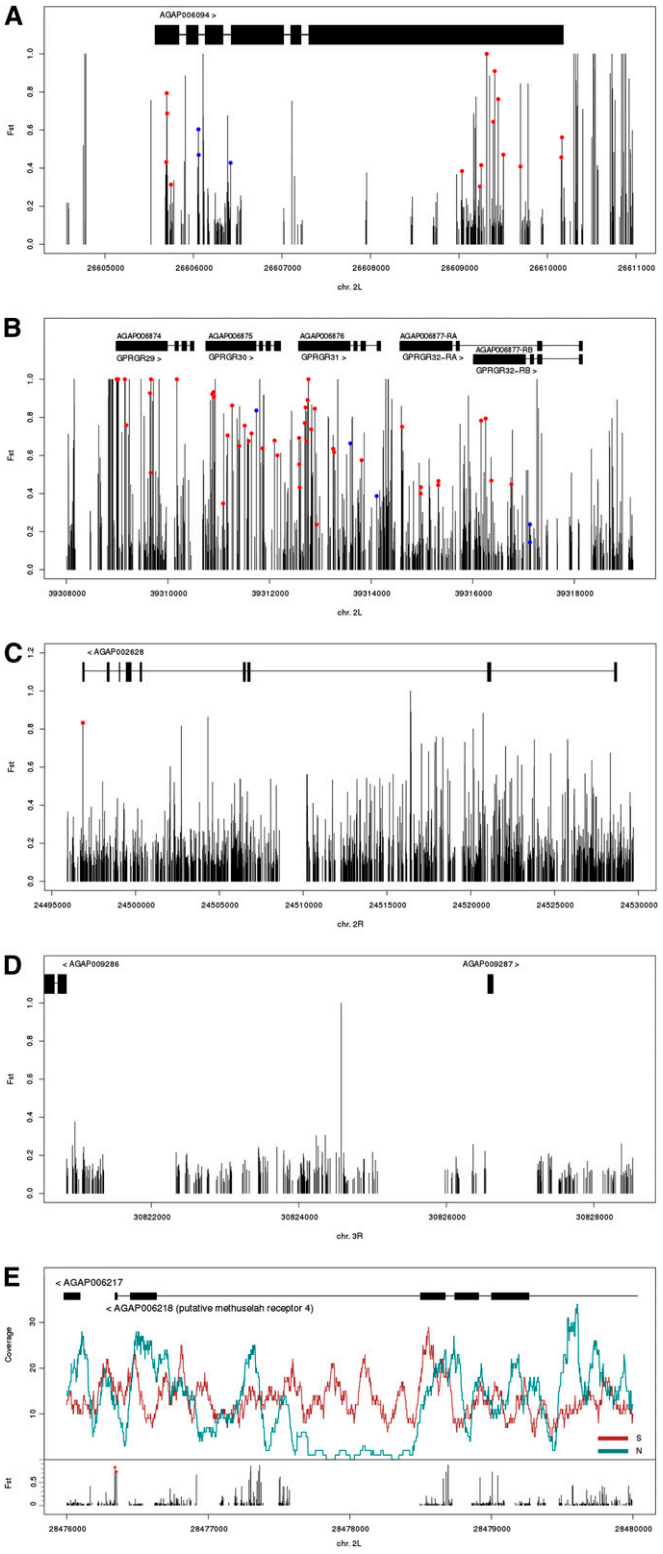

Figure 7 Elevated F_{ST} or copy-number variation between northern and southern Cameroon populations in protein coding genes. F_{ST} is plotted for individual SNPs along the gene models (solid rectangles, exons; horizontal lines, introns). Nonsynonymous and splice site mutations are indicated by red and blue circles, respectively. Divergence in (A) the *JIL-1* ortholog, AGAP006094 in the *2La* rearrangement; (B) the cluster of gustatory receptors GPRGR29-32 in the *2La* rearrangement; (C) AGAP002628 in the *2Rb* rearrangement, a putative member of the immunoglobulin superfamily; (D) an intergenic region upstream and nearby two genes in a collinear region on 3R; (E) the *Methuselah receptor 4* ortholog AGAP006218 in the *2La* rearrangement. In addition to F_{ST} at individual SNPs, read coverage is shown for each population (northern, green line; southern, red line) in relation to the gene model.

copy-number variation in the penultimate exon of this gene, a result that will require follow-up validation by PCR and sequencing. Considering the CNV gene list as a whole, we noted a significant excess of candidate genes on the X chromosome (139) relative to the autosomes ($\chi^2 = 346.7$, d.f. = 1, P-value $< 2.2 \times 10^{-16}$). If this result is largely driven by greater sequence divergence on the X chromosome between the two populations, we might have expected to see an excess of X-linked candidate genes overlapping the top 1% F_{ST} windows in the collinear genome, but we did not ($\chi^2 = 1.2$, d.f. = 1, P-value = 0.27). Also intriguing is the apparent enrichment of CNV inside the rearranged genome (114 candidate genes) relative to the collinear genome ($\chi^2 = 31.2$, d.f. = 1, P-value $= 2.35 \times 10^{-8}$), but in this case we cannot rule out sequence divergence as the driving factor.

Array genotyping vs. population resequencing

In a previous study (White *et al.* 2007a), we used a gene-based Affymetrix microarray with 25-bp probes to perform single-feature polymorphism (SFP) mapping (*e.g.*, Borevitz *et al.* 2003; Turner *et al.* 2005; Turner *et al.* 2008) of divergence between alternative arrangements of *2La* in a single population sample from Cameroon, near the center of the inversion cline. Significantly elevated divergence inside the rearrangement was mapped to two ~1.5-Mb regions near the inversion breakpoints, a finding broadly consistent with the present resequencing approach (see Figure 6, top). Nevertheless, much detail is missed with microarray-based mapping. Population resequencing revealed 25,915 total SNPs with $F_{ST} = 1$ inside *2La*, only 92 (0.36%) of which were interrogated by probes on the microarray. Moreover, of the 92 interrogated SNPs, only 8 (8.7%) were detected as differentiated SFPs on the microarray platform. Of the 84 SNPs that went undetected by the microarray, many may be true negatives in the previous population sampled due to gene flux in the center of the cline. However, 38 of the 84 SNPs fall outside the central (*i.e.*, 6–21 bp) region of the 25-bp probes on the array, and if they escaped detection for this reason, may have instead been false negatives in the earlier study.

Discussion

Population resequencing of the *An. gambiae* S form along a climatic cline in Cameroon revealed low overall genomic differentiation between populations near its endpoints, a distance of ~500 km. This is consistent with evidence from microsatellite and mtDNA markers, which indicate little or no population structure in Cameroon (Slotman *et al.* 2007)

and, more generally, only very shallow population structure across the entire African continent (Lehmann *et al.* 2003) except for that imposed by the Great Rift Valley (Lehmann *et al.* 1999). Nevertheless, polymorphic chromosomal inversions *2La* and *2Rb* are nearly fixed at the northern end and virtually absent at the southern end of the Cameroon cline. At the sequence level, alternative arrangements of both chromosomal inversions are strongly differentiated in contrast to most of the collinear genome. Compelling circumstantial evidence from polytene chromosome analysis has long suggested that these chromosomal inversions are targets of selection in *An. gambiae*, based on their frequency in relation to seasonal, latitudinal, and even microspatial gradients of aridity (Coluzzi *et al.* 1979; Rishikesh *et al.* 1985; Petrarca *et al.* 1990; Powell *et al.* 1999). By sequencing separately the alternative *2La/a* and $2L+^{a}/+^{a}$ homokaryotypes sampled from a single population near the center of the Cameroon cline, we were able to show that they are not genetically isolated. In collinear genomic regions, SNPs whose frequencies were distinctive at opposite ends of the cline are homogenized in the center, and the same occurs—although to a lesser degree—in rearranged regions of the genome through gene flux in inversion heterozygotes. Without selection acting to maintain the cline, the pattern of genetic differentiation would quickly erode. Taken together, this evidence is the strongest indication to date that spatially varying selection—not demographic history—is responsible for the clinal pattern of genetic differences in *An. gambiae* from Cameroon.

Judging from the drastically higher levels of divergence in rearranged *vs.* collinear genomic regions, our findings suggest that inversions play a disproportionate role in ecological adaptation in *An. gambiae*. This notion is not altogether surprising, when considered in the context of the eponymous *An. gambiae* complex, a group that includes *An. gambiae* and at least six other closely related and morphologically indistinguishable African sibling species (Coluzzi *et al.* 1979; White *et al.* 2011 for review). Many of these species are thought to have radiated through a process of ecological speciation, driven by larval habitat competition (Costantini *et al.* 2009; Simard *et al.* 2009). In this group, more than 120 polymorphic chromosomal inversions and 10 fixed inversions are nonrandomly distributed physically (among the five chromosome arms) and taxonomically (among the member species) (Coluzzi *et al.* 2002). Chromosome 2R contains 58% (18/31) of the common polymorphic inversions, although it represents $<$30% of the polytene (euchromatic) complement, while the X chromosome harbors 50% of the fixed inversions (5/10) despite its relatively small (11%) share of the polytene complement. Most species in the complex have relatively limited distributions and little or no inversion polymorphism. The two species that can be considered the most "successful" on the basis of their dominance across much of sub-Saharan Africa—*An. gambiae* and *An. arabiensis*—carry an abundance of polymorphic inversions on chromosome 2, although they are distinguished by five fixed inversion differences on the X chromosome (Coluzzi *et al.* 2002). The polymorphic inversions, some of which are shared through hybridization between *An. gambiae* and *An. arabiensis* (see below), are presumed to be responsible for the ecological flexibility of the two species.

Not only in the *An. gambiae* complex, but also in other major malaria vectors in the same Anopheles subgenus (*Cellia*), similar nonuniform distributions of fixed and polymorphic inversions are observed (Kitzmiller 1977; Xia *et al.* 2010; Sharakhova *et al.* 2011). The enrichment of polymorphic inversions on chromosome 2 in *An. gambiae* is observed on the homologous chromosome arms in *Anopheles stephensi* (2R, 3L) and *An. funestus* (2R, 3R) (Xia *et al.* 2010; Sharakhova *et al.* 2011). Even more remarkable, at least some of the independently derived inversions on 2R nonrandomly share common genes between species (this is not true of 2L and its homologs). The colocalization of similar sets of genes inside 2R inversions is not simply due to retention of ancestral gene order, because many of the genes have been extensively reshuffled into new gene combinations in the three lineages (Sharakhov *et al.* 2002; Sharakhova *et al.* 2011). Thus, the possibility exists that shared gene content may reflect similar ecological adaptations to common environmental pressures. *An. stephensi* is an Asian vector that does not occur in Africa. However, *An. funestus* is second only to *An. gambiae* as a major vector of malaria in Africa and is sympatric with *An. gambiae* over much of its continent-wide distribution (Coetzee and Fontenille 2004), including over the wide range of eco-climatic settings in Cameroon (Ayala *et al.* 2009). Chromosomal inversions in *An. funestus* are correlated with degree of aridity, just as they are in *An. gambiae*. Along the same latitudinal gradient in Cameroon featured in the present study of *An. gambiae*, *An. funestus* shows analogous clinal patterns of inversion frequency (Cohuet *et al.* 2005; Ayala *et al.* 2011). Two recent studies present strong evidence supporting the role of environmental selection in shaping the distribution of *An. funestus* inversions in Cameroon (Ayala *et al.* 2011; D. Ayala, R. F. Guerrero, and M. Kirkpatrick, unpublished data). The sequencing and assembly of an *An. funestus* reference genome in the framework of a larger anopheline sequencing project (Besansky 2008) will provide an unprecedented opportunity for comparative genomics of adaptation along the same environmental gradient in Cameroon.

Kolaczkowski *et al.* (2011) recently compared *D. melanogaster* populations sampled from opposite ends of the Australian climatic cline by population resequencing. Like this study, they observed lower heterozygosity at the ends of chromosome arms near telomeres and centromeres, a pattern that presumably reflects perennially reduced recombination in these regions coupled with linked selection (Begun and Aquadro 1992) (note that recombination inside inversions is reduced only in heterokaryotypes, but is normal within arrangement classes). In addition, they observed comparable levels of average genome-wide sequence divergence

between populations at the ends of the cline (mean F_{ST} based on 1-kb windows = 0.112 in *D. melanogaster* and 0.123 in *An. gambiae* excluding rearranged regions). However, a striking contrast between studies is the considerably lower F_{ST} in the region spanned by the clinal chromosomal rearrangements *In3RP* and *In2Lt* (0.129, 0.116), compared to levels estimated for *2La* and *2Rb* in *An. gambiae* (0.247, 0.149). The correspondingly smaller difference in divergence between rearranged and collinear genomic regions in *D. melanogaster* led the authors to emphasize the genome-wide distribution of candidates for environmental adaptation, whereas in *An. gambiae*, outliers of divergence are clearly concentrated in the rearranged regions. One logistical difference between studies is that the *D. melanogaster* populations were not sorted by karyotype prior to sequencing. The pooling of the predominant arrangement together with the less frequent arrangement in both population samples may factor into the lower divergence estimates for rearrangements in *Drosophila*. However, it seems unlikely that this technical consideration alone can account for the large differences, as can be seen for *2Rb*, which was not selected for but showed the expected clinal pattern.

One possible population genetic explanation for differences in the degree of genetic differentiation between collinear and rearranged regions in *D. melanogaster* and *An. gambiae* may have to do with the balance between gene flux and the strength of selection maintaining inversions in these species, although little information is available in this regard from either species. All other things being equal, the apparently younger age estimations for inversions in *Drosophila* (on the order of N_e generations) (Andolfatto *et al.* 2001) compared to the estimated ages for the *2La* and *2Rb* inversion polymorphisms in *An. gambiae* ($\sim$2.6–2.7 N_e) (White *et al.* 2007a, 2009) should have led to higher rates of divergence in rearranged regions of *Drosophila* (Feder and Nosil 2009), the opposite of what was observed. Aside from differences in the selection–migration balance, a second explanation seems at least as likely, and has its basis in hybridization and introgression between *An. gambiae* and *An. arabiensis* (Besansky *et al.* 2003). *An. arabiensis*, hypothesized to be basal in the *An. gambiae* complex phylogeny (Ayala and Coluzzi 2005), is an arid-adapted species fixed for the *2La* arrangement (and polymorphic for *2Rb*/+^{b}). *An. gambiae* is proposed to have arisen more recently in the humid rainforests of Central Africa, with a karyotype similar to present-day rainforest populations (*2L*+^{a}; *2R*+^{b}) (Ayala and Coluzzi 2005). Acquisition of the *2La* and *2Rb* arrangements by *An. gambiae* through secondary contact and hybridization with *An. arabiensis* would have allowed it to considerably expand its range into the arid savannas (Besansky *et al.* 2003; Ayala and Coluzzi 2005). The codistribution of inversions *2La* and *2Rb* in contemporary populations of *An. gambiae* and *An. arabiensis* across Africa is considered to be the product of at least one interspecific introgression event (White *et al.* 2007a, 2009). If this interpretation is correct, the relatively high levels of divergence associated with these rearrangements in *An. gambiae*, in contrast to rearrangements in *D. melanogaster*, may be due to the fact that *2La* and *2Rb* were "captured" in their entirety from a different species. The higher divergence observed between alternative arrangements of *2La* compared to *2Rb* in *An. gambiae* may also have something to do with *An. arabiensis*, in that *2La* is fixed in that species whereas *2Rb* is polymorphic and subject to gene flux. How often ecological adaptation is aided by interspecific transfer of inversions remains to be seen, but a conceptually related process of introgression between Mexican and North American populations of *Rhagoletis pomonella* may have facilitated the host shift from hawthorn to apple (Feder *et al.* 2003).

The Australian populations of *Drosophila* studied by Kolaczkowski *et al.* (2011) were tropical in the north and temperate in the south, whereas the Cameroon populations of *An. gambiae* were all tropical, although spanning a steep gradient of aridity. Although the precise combination of selective agents could differ between these two examples, temperature, humidity, and rainfall appear to be important common factors that vary along both clines (Umina *et al.* 2005). As such, there is an intriguing correspondence of functional themes among candidate genes identified in the two studies, and in some cases orthologous genes are implicated. For example, genes that function in signaling pathways were enriched in both studies. Candidate genes functioning in *EGFR*, *TGF-β*, and *EcR* pathways could influence clinal variation in body size, metabolism, developmental processes, and other life-history traits (Kolaczkowski *et al.* 2011). *Anopheles* orthologs of several *Drosophila* candidate genes in this category were identified. Another striking commonality was the large number of gustatory receptors, ionotropic receptors, and ion-channel-related genes, including cyclic nucleotide-gated channels, implicated in chemo- and thermo-sensation. Finally, both studies identified a number of the same genes implicated in regulation of chromatin and transcription. It will be of great interest to learn if any of these shared candidate *Drosophila* and *Anopheles* genes represent parallel adaptive responses to similar environmental selective pressures. Reverse population genomic analyses such as these are powerful for generating hypotheses about the adaptive process. Yet the daunting challenge remains to directly connect these specific candidate genes to traits that vary clinally. Our finding of substantial amounts of gene flux at the center of the cline in Cameroon suggests that one particularly promising approach to infer adaptation at specific loci would be to perform genome-wide association studies at middle latitudes, to test whether SNP variants are associated with traits already implicated in climatic adaptation, such as thermal and desiccation resistance (Gray *et al.* 2009; Rocca *et al.* 2009).

Climate strongly influences the genetic constitution of several *Drosophila* species, such that adaptive polymorphisms (chromosomal inversions) associated with climate have been proposed as tools for monitoring climate change

(Umina *et al.* 2005; Balanya *et al.* 2009). Similar associations of adaptive polymorphisms and climate exist in anopheline vectors of malaria. However, in the case of these human disease vectors, the adaptive polymorphisms that facilitate survival in otherwise inhospitable territory have public health as well as evolutionary importance, because expansion of the species range or seasonal activity is accompanied by increased malaria transmission. An understanding of the mechanisms underlying adaptation to climate can potentially provide a much-needed new arsenal of tools targeted against vectors.

Acknowledgments

We thank D. Ayala for producing Figure 1; I. Lanc and S. Emrich for programming to execute SHRiMP by batch processing via Condor; D. Thain for assistance with storage of large data files; J. Ford for sample pooling and sequencing; the entomology staff at OCEAC for expert technical assistance; and D. Schrider for early contributions to the analysis. C.C. was supported by a Fellowship from the Eck Institute for Global Health and the University of Notre Dame. Funding was provided by the National Institutes of Health grant R01AI076584 (to N.J.B.).

Literature Cited

Akey, J. M., A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly *et al.*, 2010 Tracking footprints of artificial selection in the dog genome. Proc. Natl. Acad. Sci. USA 107: 1160–1165.

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

Ayala, D., C. Costantini, K. Ose, G. C. Kamdem, C. Antonio-Nkondjio *et al.*, 2009 Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar. J. 8: 307.

Ayala, D., M. C. Fontaine, A. Cohuet, D. Fontenille, R. Vitalis *et al.*, 2011 Chromosomal inversions, natural selection and adaptation in the malaria vector Anopheles funestus. Mol. Biol. Evol. 28: 745–758.

Ayala, F. J., and M. Coluzzi, 2005 Chromosome speciation: Humans, Drosophila, and mosquitoes. Proc. Natl. Acad. Sci. USA 102(Suppl 1): 6535–6542.

Baker, D. A., T. Nolan, B. Fischer, A. Pinder, A. Crisanti *et al.*, 2011 A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. BMC Genomics 12: 296.

Balanya, J., R. B. Huey, G. W. Gilchrist, and L. Serra, 2009 The chromosomal polymorphism of *Drosophila subobscura*: a microevolutionary weapon to monitor global change. Heredity 103: 364–367.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356: 519–520.

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics 134: 869–893.

Besansky, N. J., 2008 Genome analysis of vectorial capacity in major Anopheles vectors of malaria parasites, http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/AnophelesGenomesProposal_Aug3.pdf.

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.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon, 2006 Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Mol. Biol. Evol. 23: 773–783.

Borevitz, J. O., D. Liang, D. Plouffe, H. S. Chang, T. Zhu *et al.*, 2003 Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res. 13: 513–523.

Coetzee, M., and D. Fontenille, 2004 Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. Insect Biochem. Mol. Biol. 34: 599–605.

Cohuet, A., I. Dia, F. Simard, M. Raymond, F. Rousset *et al.*, 2005 Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. Genetics 169: 301–311.

Collins, F. H., M. A. Mendez, M. O. Rasmussen, P. C. Mehaffey, N. J. Besansky *et al.*, 1987 A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. Am. J. Trop. Med. Hyg. 37: 37–41.

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

Coluzzi, M., A. Sabatini, A. Della Torre, M. A. Di Deco, and V. Petrarca, 2002 A polytene chromosome analysis of the *Anopheles gambiae* species complex. Science 298: 1415–1418.

Costantini, C., D. Ayala, W. M. Guelbeogo, M. Pombi, C. Y. Some *et al.*, 2009 Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. BMC Ecol. 9: 16.

De Jong, G., and Z. Bochdanovits, 2003 Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. J. Genet. 82: 207–223.

Dobzhansky, T., 1947 Adaptive changes induced by natural selection in wild populations of Drosophila. Evolution 1: 1–16.

Ellison, C. E., C. Hall, D. Kowbel, J. Welch, R. B. Brem *et al.*, 2011 Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc. Natl. Acad. Sci. USA 108: 2831–2836.

Endler, J. A., 1977 Geographic variation, speciation, and clines. Monogr. Popul. Biol. 10: 1–246.

Feder, J. L., and P. Nosil, 2009 Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? Evolution 63: 3061–3075.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith *et al.*, 2003 Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proc. Natl. Acad. Sci. USA 100: 10314–10319.

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

Gibbs, A. G., 2002 Water balance in desert Drosophila: lessons from non-charismatic microfauna. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 133: 781–789.

Gillies, M. T., and B. De Meillon, 1968 *The Anophelinae of Africa South of the Sahara*. South African Institute for Medical Research, Johannesburg, South Africa.

Gray, E. M., K. A. Rocca, C. Costantini, and N. J. Besansky, 2009 Inversion 2La is associated with enhanced desiccation resistance in *Anopheles gambiae*. Malar. J. 8: 215.

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.

Hoffmann, A. A., and A. R. Weeks, 2007 Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. Genetica 129: 133–147.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.

Holt, K. E., Y. Y. Teo, H. Li, S. Nair, G. Dougan *et al.*, 2009 Detecting SNPs and estimating allele frequencies in clonal bacterial populations by sequencing pooled DNA. Bioinformatics 25: 2074–2075.

Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab *et al.*, 2002 The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298: 129–149.

Hosack, D. A., G. Dennis Jr. B. T. Sherman, H. C. Lane, and R. A. Lempicki, 2003 Identifying biological themes within lists of genes with EASE. Genome Biol. 4: R70.

Huang, D. W., B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord *et al.*, 2007 The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8: R183.

Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44–57.

Hudson, R. R., M. Slatkin, and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583–589.

Kennington, W. J., J. Gockel, and L. Partridge, 2003 Testing for asymmetrical gene flow in a *Drosophila melanogaster* body-size cline. Genetics 165: 667–673.

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

Kitzmiller, J. B., 1977 Chromosomal differences among species of Anopheles mosquitoes. Mosquito Systematics 9: 112–122.

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 Introduction, pp. 1–52 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas, and J. R. Powell. CRC Press, Boca Raton, FL.

Laayouni, H., E. Hasson, M. Santos, and A. Fontdevila, 2003 The evolutionary history of *Drosophila buzzatii*, XXXV: inversion polymorphism and nucleotide variability in different regions of the second chromosome. Mol. Biol. Evol. 20: 931–944.

Lawson, D., P. Arensburger, P. Atkinson, N. J. Besansky, R. V. Bruggner *et al.*, 2009 VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res. 37: D583–D587.

Lee, Y., A. J. Cornel, C. R. Meneses, A. Fofana, A. G. Andrianarivo *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*. Malar. J. 8: 75.

Lehmann, T., W. A. Hawley, H. Grebert, M. Danga, F. Atieli *et al.*, 1999 The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. J. Hered. 90: 613–621.

Lehmann, T., M. Licht, N. Elissa, B. T. Maega, J. M. Chimumbwa *et al.*, 2003 Population structure of *Anopheles gambiae* in Africa. J. Hered. 94: 133–147.

Li, H., J. Ruan, and R. Durbin, 2008 Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18: 1851–1858.

Li, Y. F., J. C. Costello, A. K. Holloway, and M. W. Hahn, 2008 "Reverse ecology" and the power of population genomics. Evolution 62: 2984–2994.

Lobo, N. F., D. M. Sangare, A. A. Regier, K. R. Reidenbach, D. A. Bretz *et al.*, 2010 Breakpoint structure of the *Anopheles gambiae* 2Rb chromosomal inversion. Malar. J. 9: 293.

Machado, C. A., T. S. Haselkorn, and M. A. Noor, 2007 Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics 175: 1289–1306.

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.

Petrarca, V., G. Sabatinelli, M. A. Di Deco, and M. Papakay, 1990 The *Anopheles gambiae* complex in the Federal Islamic Republic of Comoros (Indian Ocean): some cytogenetic and biometric data. Parassitologia 32: 371–380.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, Oxford.

Powell, J. R., V. Petrarca, A. della Torre, A. Caccone, and M. Coluzzi, 1999 Population structure, speciation, and introgression in the *Anopheles gambiae* complex. Parassitologia 41: 101–113.

Regnard, C., T. Straub, A. Mitterweger, I. K. Dahlsveen, V. Fabian *et al.*, 2011 Global analysis of the relationship between JIL-1 kinase and transcription. PLoS Genet. 7: e1001327.

Rishikesh, N., M. A. Di Deco, V. Petrarca, and M. Coluzzi, 1985 Seasonal variations in indoor resting *Anopheles gambiae* and *Anopheles arabiensis* in Kaduna, Nigeria. Acta Trop. 42: 165–170.

Rocca, K. A., E. M. Gray, C. Costantini, and N. J. Besansky, 2009 2La chromosomal inversion enhances thermal tolerance of *Anopheles gambiae* larvae. Malar. J. 8: 147.

Rumble, S. M., P. Lacroute, A. V. Dalca, M. Fiume, A. Sidow *et al.*, 2009 SHRiMP: accurate mapping of short color-space reads. PLOS Comput. Biol. 5: e1000386.

Santolamazza, F., A. Della Torre, and A. Caccone, 2004 Short report: a new polymerase chain reaction-restriction fragment length polymorphism method to identify *Anopheles arabiensis* from *An. gambiae* and its two molecular forms from degraded DNA templates or museum samples. Am. J. Trop. Med. Hyg. 70: 604–606.

Schaeffer, S. W., 2008 Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. Evolution 62: 3082–3099.

Schaeffer, S. W., and W. W. Anderson, 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics 171: 1729–1739.

Sharakhov, I. V., A. C. Serazin, O. G. Grushko, A. Dana, N. Lobo *et al.*, 2002 Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. Science 298: 182–185.

Sharakhova, M. V., P. George, I. V. Brusentsova, S. C. Leman, J. A. Bailey *et al.*, 2010 Genome mapping and characterization of the *Anopheles gambiae* heterochromatin. BMC Genomics 11: 459.

Sharakhova, M. V., A. Xia, S. C. Leman, and I. V. Sharakhov, 2011 Arm-specific dynamics of chromosome evolution in malaria mosquitoes. BMC Evol. Biol. 11: 91.

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

Slotman, M. A., F. Tripet, A. J. Cornel, C. R. Meneses, Y. Lee *et al.*, 2007 Evidence for subdivision within the M molecular form of *Anopheles gambiae*. Mol. Ecol. 16: 639–649.

Storz, J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14: 671–688.

Toure, Y. T., V. Petrarca, S. F. Traore, A. Coulibaly, H. M. Maiga *et al.*, 1998 The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. Parassitologia 40: 477–511.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3: e285.

Turner, T. L., M. T. Levine, M. L. Eckert, and D. J. Begun, 2008 Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. Genetics 179: 455–473.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin, 2010 Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. 42: 260–263.

Umina, P. A., A. R. Weeks, M. R. Kearney, S. W. McKechnie, and A. A. Hoffmann, 2005 A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science 308: 691–693.

Vogel, C., S. A. Teichmann, and C. Chothia, 2003 The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. Development 130: 6317–6328.

White, B. J., M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo *et al.*, 2007a Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. PLoS Genet. 3: e217.

White, B. J., F. Santolamazza, L. Kamau, M. Pombi, O. Grushko *et al.*, 2007b Molecular karyotyping of the 2La inversion in *Anopheles gambiae*. Am. J. Trop. Med. Hyg. 76: 334–339.

White, B. J., C. Cheng, D. Sangare, N. F. Lobo, F. H. Collins *et al.*, 2009 The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. Genetics 183: 275–288.

White, B. J., F. H. Collins, and N. J. Besansky, 2011 Evolution of *Anopheles gambiae* in relation to humans and malaria. Annu. Rev. Ecol. Evol. Syst. 42: (in press).

Xia, A., M. V. Sharakhova, S. C. Leman, Z. Tu, J. A. Bailey *et al.*, 2010 Genome landscape and evolutionary plasticity of chromosomes in malaria mosquitoes. PLoS ONE 5: e10592.

Communicating editor: M. Long

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2011/12/30/genetics.111.137794.DC1

Ecological Genomics of *Anopheles gambiae* Along a Latitudinal Cline: A Population-Resequencing Approach

Changde Cheng, Bradley J. White, Colince Kamdem, Keithanne Mockaitis, Carlo Costantini, Matthew W. Hahn, and Nora J. Besansky

DOI: 10.1534/genetics.111.137794

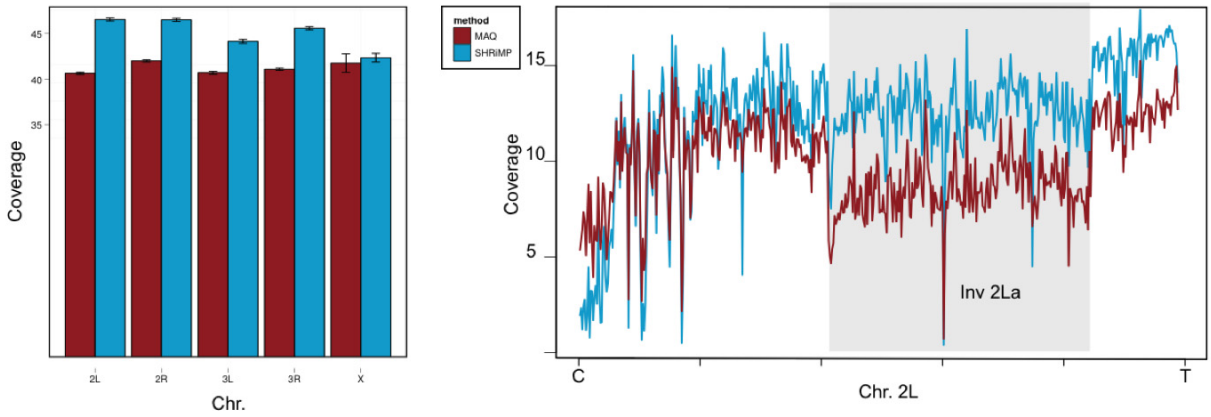

Figure S1 Read coverage from uniquely m pped reads based on MAQ or SHRiMP read alignment programs. Left panel, average read coverage by chromosome arm for the full data set. Right panel, average read coverage on chromosome 2L for the northern *2La*/a population (mapped to the $2L+^{a}/+^{a}$ *An. gambiae* PEST reference), plotted in 10-kb non-overlapping windows.

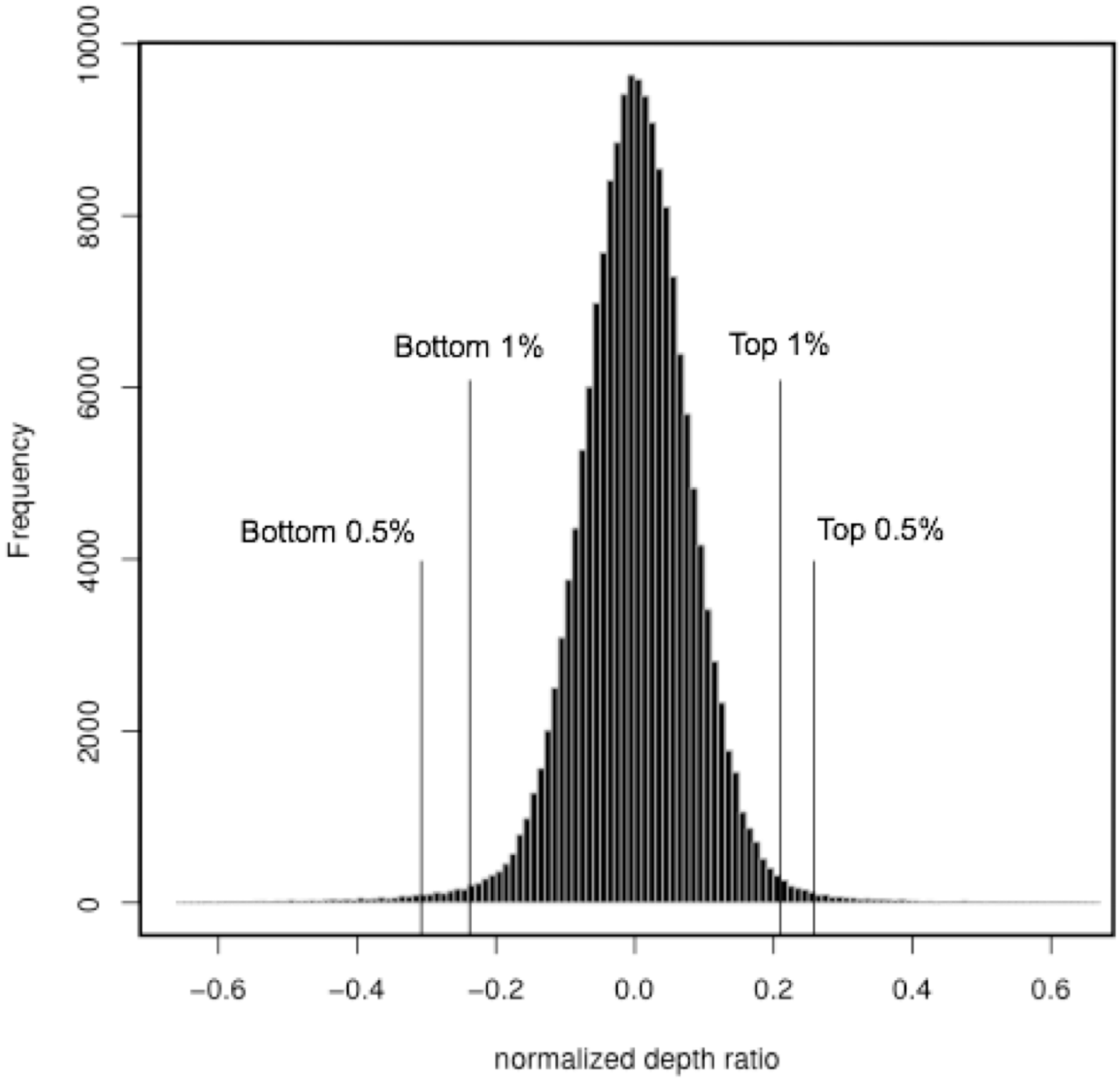

Figure S2 Histogram of normalized read depth ratios measured in non-overlapping 1-kb genomic windows between northern and southern populations. Tail cut-offs of 0.5% were used to infer copy number variation.

Table S1 Mean F_{ST} based on 1-kb windows for population pairs across the collinear genome (and in the region spanning rearrangements)

	South (*2L+*^{a}/^{a})	Central (*2La*/*a*)	Central (*2L+*^{a}/^{a})
South (*2L+*^{a}/^{a})	---	---	---
Central (*2La*/*a*)	0.127 (0.137)	---	---
Central (*2L+*^{a}/^{a})	0.119 (0.119)	0.114 (0.121)	---
North (*2La*/*a*)	0.123 (0.135)	0.117 (0.118)	0.110 (0.120)

Table S2 Mean nucleotide diversity (π) and divergence (F_{ST}) for northern and southern population samples of *An. gambiae* S form based on 1-kb windows across the euchromatic chromosome arms

Chromosome	π collinear (rearranged, flanking[1])		F_{ST} collinear (rearranged)
	North	South	
2L	0.014 (0.014, 0.016)	0.011 (0.012, 0.013)	0.124 (0.247)
2R	0.014 (0.016, 0.016)	0.012 (0.013, 0.013)	0.122 (0.149)
3L	0.014	0.012	0.123
3R	0.015	0.012	0.123
	0.011	0.008	0.121

[1]Flanking, average diversity across 5-M flanking the rearrangement on centromeric and telomeric sides

Table S3 ***Anopheles gambiae*** **(A. gam) genes overlapping the most diverged (top 1% F_{ST}) 1-kb windows between northern and southern populations inside the 2La rearrangement, and their** ***D. melanogaster*** **(D. mel) homologs**

A. gam Gene ID	D. mel FlyBase ID	Homology Type	Description (Gene Name)	GO Biological ocess	MozAtlas[1]	DAVID cluster[2]
AGAP005781	FBgn0036208	1-to-1	CG10361	cellular amino acid metabolic process	F	
AGAP005784	FBgn0020623	1-to-1	eptidyl-alpha-hydroxyglycine-alpha-amidating lyase	-	male Acp	
AGAP005790	FBgn0043470		lambdaTry	proteolysis		
AGAP005796	FBgn0003041	1-to-1	pebble	cell adhesion	+	
AGAP005799	FBgn0011577	1-to-1	division abnormally delayed	decapentaplegic signaling pathway		
AGAP005806	FBgn0044328	1-to-1	CG32052	-	+	
AGAP005807	FBgn0053205	1-to-many	CG33205	mesoderm development		
	FBgn0052050	1-to-many		-		
AGAP005816	FBgn0035756	1-to-1	unc-13-4A	neurotransmitter secretion		
AGAP005817	FBgn0016131	1-to-1	Cyclin-dependent kinase 4	protein phosphorylation	male testis	2
AGAP005898	FBgn0011817	1-to-1	nemo	negative regulation of Wnt ecepto signaling pathway	male testis	2
AGAP005920	FBgn0029094	1-to-1	anti-silencing factor 1	nucleosome assembly	F	
AGAP005924	FBgn0035381	1-to-1	CG9965			
AGAP005928	FBgn0035981	1-to-1	CG4452	-		
AGAP005964	FBgn0052423	1-to-1	alan shepard	gravitaxis		2
AGAP006026	FBgn0028431		glutamate	-		3

			receptor IB			
AGAP006058	FBgn0003089	apparent_1-to-1	pipe	Toll signaling pathway		
AGAP006089	FBgn0052062	1-to-1	Ataxin-2 binding protein 1	nervous system development		2
AGAP006094	FBgn0020412	1-to-1	JIL-1	chromatin organization		2
AGAP006156	FBgn0085401	1-to-1	CG34372	G-protein coupled receptor protein signaling pathway		2,3,4
AGAP006165	FBgn0020306		domino	cell proliferation		
AGAP006274	FBgn0085414	1-to-1	dpr12	sensory perception of chemical stimulus		
AGAP006277						
AGAP006282	FBgn0259878	1-to-1	follistatin	negative egulation of activin receptor signaling pathway		
AGAP006330	FBgn0042185	1-to-1	CG18769	-		
AGAP006347	FBgn0033494	1-to-1	CNQ potassium channel	potassium ion transport	female salivary	3,4
AGAP006349	FBgn0032449	1-to-1	CG17036	transport	F	
AGAP006381	FBgn0259744	1-to-1	CG42377	-		
AGAP006452	FBgn0010482	1-to-1	lethal (2) 01289	cell redox homeostasis		
AGAP006516	FBgn0260660	1-to-1	multiplexin	motor axon guidance	+	1
AGAP006590	FBgn0085447	apparent_1-to-1	still life	regulation of axonogenesis		
AGAP006633	FBgn0034135	1-to-1	Syntrophin-like 2	egulation of synaptic growth at	male testis	

				neuromuscular junction		
AGAP006715	FBgn0023083	1-to-1	frayed	hitin-based embryonic uticle iosynthetic process	+	2
AGAP006754						
AGAP006762	FBgn0085390	1-to-1	Dgk	phosphorylation	F	
AGAP006824	FBgn0085453	1-to-1	CG34424	-	m le M lpighian	2
AGAP006843	FBgn0036879	m ny-to-many	uticular protein 76Bb	-		1
	FBgn0036878	m ny-to-many		-		
AGAP006848	FBgn0036879	m ny-to-many	uticular protein 76Bb	-		1
	FBgn0036878	m ny-to-many		-		
AGAP006849	FBgn0036879	m ny-to-many	uticular protein 76Bb	-		1
	FBgn0036878	m ny-to-many		-		
AGAP006852	FBgn0036879	m ny-to-many	uticular protein 76Bb	-		1
	FBgn0036878	m ny-to-many		-		
AGAP006853	FBgn0036879	m ny-to-many	uticular protein 76Bb	-		1
	FBgn0036878	m ny-to-many		-		
AGAP006872	FBgn0045980	1-to-1	nimA-like kinase	protein phosphorylation		2
AGAP006874						3,4
AGAP006875						3,4
AGAP006876						3,4

AGAP006884	FBgn0034854		CG3493	protein targeting to Golgi	+	
AGAP006935	FBgn0041775	1-to-1	trailer hitch	E to Golgi vesicle-mediated transport		
AGAP007005	FBgn0050089	1-to-1	CG30089	-	m le testis	
AGAP007008	FBgn0259145	1-to-1	CG42260	transmembrane transport		3 4
AGAP007029	FBgn0040259	m ny-to-many	gt86Da	metabolic process	male midgut	
	FBgn0040251	m ny-to-many		metabolic process		
	FBgn0026314	m ny-to-many		metabolic process		
	FBgn0040257	m ny-to-many		metabolic process		
	FBgn0026315	m ny-to-many		metabolic process		
	FBgn0051002	m ny-to-many		metabolic process		
	FBgn0040255	m ny-to-many		metabolic process		
	FBgn0040253	m ny-to-many		metabolic process		
	FBgn0040252	m ny-to-many		metabolic process		
	FBgn0040250	m ny-to-many		metabolic process		
	FBgn0039087	m ny-to-many		metabolic process		
	FBgn0039086	m ny-to-many		metabolic process		
	FBgn0039085	m ny-to-many		metabolic process		
AGAP007030	FBgn0039728		CG7896	-		
AGAP007046	FBgn0005564	1-to-1	shaker cognate	potassium ion	+	3 4

			l	transport	
AGAP007049	FBgn0034638	1-to-1	CG10433	defense response	M

[1]Pattern of gene expression from MozAtlas (www.tissue-atl s.org): +, at least one probe detected as expressed; F/M, gender-iased expression in fem les (F) or males (M). Cases where expression is unique to one sex and one tissue are explicitly noted.
[2] ee Table 1.

Table S4 ***Anopheles gambiae*** **(A. gam) genes overlapping the most diverged (top 1% F_{ST}) 1-kb windows between northern and southern populations inside the 2Rb rearrangement, and their** ***D. melanogaster*** **(D. mel) homologs**

A. gam Gene D	D. mel FlyBase ID	Homology Type	Description (Gene ame)	GO Biological ocess	MozAtlas[1]	DAVID Cluster[2]
AGAP002300	FBgn0038269	1-to-1	Rrp6	mRNA polyadenylation	F	
AGAP002303	FBgn0030710	1-to-1	CG8924	-		
AGAP002305	FBgn0039558	1-to-1	CG4980	nuclear mRNA splicing, via spliceosome	+	
AGAP002310	FBgn0039709	1-to-1	Cad99C, isoform B	calcium-dependent cell-cell adhesion		
AGAP002315	FBgn0039705	1-to-1	CG31033, isoform C	-	+	
AGAP002325	FBgn0017590	1-to-1	klingon	R7 cell differentiation		5
AGAP002372	FBgn0003267	1-to-1	rough, isoform B	regulation of transcription, DNA-dependent		
AGAP002374	FBgn0004237	1-to-1	heterogeneous nuclear ribonucleoprotein at 87F, isoform B	regulation of alternative nuclear mR A splicing, via spliceosome	F	
AGAP002375	FBgn0038903	1-to-1	Rp 12	transcription from RNA polymerase I promoter	F	
AGAP002376	FBgn0039255	1-to-1	CG13646	amino acid transmembrane transport	+	
AGAP002443	FBgn0036260		rhodopsin 7	G-protein coupled receptor protein signaling pathway		
AGAP002445						
AGAP002590						
AGAP002592	FBgn0053126		neural lazarillo	response to oxidative stress	F	
AGAP002593	FBgn0053126		neural lazarillo	response to	F	

				oxidative stress		
AGAP002594	FBgn0053126		neural lazarillo	response to oxidative stress	F	
AGAP002628	FBgn0085382	1-to-1	CG34353	-		5
AGAP002636						
AGAP002674	FBgn0086368	1-to-1	protein O-mannosyltransferase 2	protein O-linked mannosylation		
AGAP002677	FBgn0083946	1-to-1	CG34110	sperm motility	m le testis	
AGAP002707	FBgn0030723	1-to-many	dpr18	sensory perception of chemical stimulus	+	5
AGAP002737	FBgn0031273	Apparent 1-to-1	Stretchin-Mlck	-	+	5
AGAP013321	FBgn0038929	1-to-many	CG13408	-		
AGAP013395						

[1]Pattern of gene expression from MozAtlas (www.tissue-atl s.org): +, at least one probe detected as expressed; F/M, gender-biased expression in fem les (F) or males (M). Cases where expression is unique to one sex and one tissue are explicitly noted.
[2] ee Table 1.

Table S5 ***Anopheles gambiae*** **(A. gam) genes overlapping the most diverged (top 1% F_{ST}) 1-kb windows between northern and southern populations in the ollinear genom , and their** ***D. melanogaster*** **(D. mel) homologs**

A. gam Gene ID	Chr	D. mel FlyBase ID	Homology	Description (Gene Name)	G Biological rocess	MozAtlas[1]	DAVID cluster[2]
AGAP000080	X	FBgn0004652	1-to-1	fruitless	male courtship behavior, veined wing vibration		
AGAP000179	X	FBgn0004901	1-to-many	phosphoribosylamidotransferase	purine base biosynthetic process	F	
AGAP000179	X	FBgn0041194	1-to-many		de novo' IMP biosynthetic process		
AGAP000277	X	FBgn0029995	1-to-1	CG2256	-		
AGAP000316	X	FBgn0030759	1-to-1	CG13014	-		
AGAP000320	X	FBgn0003654	1-to-many	short wing	eye photoreceptor cell differentiation	F	
		FBgn0053499	1-to-many		microtubule-based movement		
		FBgn0053497	1-to-many		microtubule-based movement		
		FBgn0052823	1-to-many		microtubule-based movement		
		FBgn0067861	1-to-many		microtubule-based movement		
AGAP000346	X	FBgn0003495		spatzle	Toll signaling pathway		
AGAP000351	X	FBgn0004842	1-to-many	neuropeptide Y receptor-like	G-protein coupled receptor protein signaling pathway		6,9
AGAP000359	X	FBgn0052499	1-to-1	chitin deacetylase-like 4	neurogenesis	+	
AGAP000399	X	FBgn0086897	1-to-many	squid	-	F	
AGAP000410	X	FBgn0041210	1-to-1	HDAC4	regulation of transcription, DNA-dependent	+	
AGAP000454	X	FBgn0086712	1-to-1	enigma	compound eye	+	

					morphogenesis		
AGAP000457	X	FBgn0086768	1-to-1	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	protein repair	+	
AGAP000484	X	FBgn0002941	1-to-1	slouch	muscle cell fate determination		8
AGAP000529	X	FBgn0261241	1-to-1	vacuolar protein sorting 16A	autophagic vacuole fusion	F	
AGAP000539	X	FBgn0030815	1-to-1	CG8945	proteolysis		
AGAP000543	X	FBgn0053544	1-to-1	Vitamin-K epoxide reductase	vitamin metabolic process	+	
AGAP000559	X						
AGAP000576	X	FBgn0022153	1-to-1	lethal (2) k05819	-	+	
AGAP000585	X	FBgn0263115		maternal gene required for meiosis	female meiosis		
AGAP000587	X	FBgn0030532	1-to-1	CG11071	-		
AGAP000598	X	FBgn0052500	1-to-many	CG32500	iron-sulfur cluster assembly		
		FBgn0052857	1-to-many		-		
		FBgn0053502	1-to-many		-		
AGAP000606	X	FBgn0038653	1-to-1	CG18208	G-protein coupled receptor protein signaling pathway		6
AGAP000620	X	FBgn0029808	1-to-1	CG42699	-		
AGAP000639	X	FBgn0043799	1-to-1	CG31381	tRNA modification	F	
AGAP000641	X					F	
AGAP000642	X					F	
AGAP000653	X	FBgn0030890	1-to-1	CG7536	-	+	6
AGAP000684	X	FBgn0004395	1-to-1	unkempt	bristle morphogenesis	+	
AGAP000686	X	FBgn0260933	1-to-1	reduced mechanoreceptor potential A	sensory perception of sound		
AGAP000708	X	FBgn0030087	1-to-1	CG7766	glycogen metabolic process		

AGAP000713	FBgn0052666	1-to-1	CG32666	protein phosphorylation	M	
AGAP000715	FBgn0029962	1-to-1	CG1402	one-carbon metabolic process		
AGAP000717	FBgn0031011	1-to-1	CG8034	transmembrane transport	M	
AGAP000718	FBgn0031002	1-to-many	CG14196	transmembrane transport	F	
	FBgn0259834	1-to-m ny		programmed cell death		
	FBgn0031012	1-to-m ny		tr nsmembr ne transport		
AGAP000743					+	
AGAP000748	FBgn0002789		muscle protein 20	ell adhesion		
AGAP000750	FBgn0037174	1-to-many	CG14457	-	+	
	FBgn0039483	1-to-many		-		
	FBgn0039485	1-to-many		-		
AGAP000810	FBgn0038816	1-to-1	Leucine-rich repeat kinase	syn pse organization	fem le M lpighian	
AGAP000821	FBgn0052816	1-to-1	CG32816	-	F	
AGAP000843	FBgn0039360	1-to-1	CG4774	phospholipid iosynthetic process	+	
AGAP000844	FBgn0053203	1-to-1	CG33203	l teral inhibition		6
AGAP000901	FBgn0030478	1-to-1	CG1640	-	M	
AGAP000915	FBgn0039816	1-to-1	CG11317	neurogenesis		
AGAP000926	FBgn0027279	1-to-1	lethal (1) G0196	-	m le Acp	
AGAP000939	FBgn0025809	1-to-1	Platelet-activating factor acetylhydrolase lpha	-	F	
AGAP000955	FBgn0039770	1-to-1	CG15537	hormone metabolic process		
AGAP000963	FBgn0031146	1-to-1	CG15449	-		
AGAP000974	FBgn0259110	1-to-1	mind-meld	proteolysis		

AGAP000977	X	FBgn0032822	1-to-1	CG10466	nuclear mRNA splicing, via spliceosome	+	
AGAP001100	2						
AGAP001124	2	FBgn0032287	1-to-1	CG6415	glycine catabolic process	F	
AGAP001125	2						6
AGAP001255	2	FBgn0027375	1-to-1	homolog of RecQ	negative regulation of mitotic cell cycle		
AGAP001358	2	FBgn0037705	1-to-1	murashka	learning or memory	+	
AGAP001390	2	FBgn0038098		CG7381	-	M	
AGAP001485	2	FBgn0052645	1-to-1	CG32645	-	+	
AGAP001486	2	FBgn0250871	1-to-1	papillote	apposition of dorsal and ventral imaginal disc-derived wing surfaces		
AGAP001488	2	FBgn0027108	1-to-1	innexin 2	intercellular transport	male es is	
AGAP001577	2	FBgn0038197	apparent_1-to-1	forkhead box, sub-group O	negative regulation of multicellular organism growth	F	
AGAP001579	2						
AGAP001721	2	FBgn0028327	1-to-1	lethal (1) G0320	-	F	6
AGAP001722	2	FBgn0039084		CG10175	metabolic process		
AGAP001730	2	FBgn0039858	1-to-1	cyclin G	negative regulation of G1/S transition of mitotic cell cycle	M	
AGAP001783	2	FBgn0040208	1-to-1	katanin 60	microtubule severing	female salivary	
AGAP001811	2	FBgn0037501	1-to-many	ionotropic receptor 84a	ionotropic glutamate receptor signaling	male es is	6

					pathway		
AGAP001935	2	FBgn0259100	1-to-1	CG42248	-		
AGAP001961	2	FBgn0039396	1-to-many	cardioacceleratory peptide receptor	G-protein coupled receptor protein signaling pathway		6,9
AGAP001962	2	FBgn0039396	1-to-many	cardioacceleratory peptide receptor	G-protein coupled receptor protein signaling pathway	M	6,9
AGAP001963	2						
AGAP002101	2	FBgn0027086	1-to-1	Isoleucyl-tRNA synthetase	isoleucyl-tRNA aminoacylation	F	
AGAP002104	2	FBgn0086604	1-to-1	CG12484	l teral inhibition	+	
AGAP002116	2	FBgn0261379	1-to-1	radish	phospholipid metabolic process	male testis	
AGAP002119	2	FBgn0259168	1-to-1	minibrain	protein phosphorylation	male testis	
AGAP002159	2	FBgn0032129	1-to-1	junctophilin	-	M	
AGAP002169	2	FBgn0015374	1-to-1	courtless	m le courtship ehavior		
AGAP002178	2	FBgn0001235	pparent_1-to-1	homothorax	peripheral nervous system development	M	8
AGAP002202	2R	FBgn0037601	many-to-many	Cyp4g15	oxidation-reduction process		
		FBgn0038076	many-to-many		o id tion-reduction process		
		FBgn0030369	many-to-many		o id tion-reduction process		
		FBgn0038236	many-to-many		insecticide metabolic process		
		FBgn0038007	many-to-many		o id tion-reduction process		
		FBgn0038006	many-to-many		o id tion-reduction process		
		FBgn0038005	many-to-many		o id tion-reduction process		

AGAP002757	2	FBgn0037720	1-to-1	CG8312	-		
AGAP002760	2	FBgn0002878	1-to-1	mutagen-sensitive 101	mitosis	male testis	
AGAP002761	2	FBgn0042083	1-to-1	CG3267	regulation of eclosion	F	
AGAP002773	2	FBgn0003499	1-to-1	stripe	central nervous system development		
AGAP002783	2	FBgn0040251		Ugt86Di	metabolic process	+	
AGAP002793	2	FBgn0003425	1-to-1	lit	glial cell migration		
AGAP002817	2	FBgn0040696	1-to-1	CG18675	-	+	
AGAP002820	2	FBgn0003710	1-to-1	temperature-induced paralytic E	sodium ion transport	female ovary	
AGAP002842	2R	FBgn0030051	1-to-many	snake	proteolysis		
		FBgn0037222	1-to-many		proteolysis		
		FBgn0038113	1-to-many		proteolysis		
		FBgn0038114	1-to-many		proteolysis		
AGAP002855	2	FBgn0262595		CG33724	-		
AGAP002856	2	FBgn0262595		CG33724	-		
AGAP002858	2	FBgn0002921	1-to-1	Na pump alpha subunit	embryonic development via the syncytial blastoderm		6
AGAP002859	2	FBgn0013995	apparent_1-to-1	Na/Ca-exchange protein	phototransduction		6
AGAP002881	2	FBgn0038874	1-to-1	ETHR	G-protein coupled receptor protein signaling pathway		6
AGAP002896	2	FBgn0261552	1-to-1	pasilla	nuclear m A splicing, via spliceosome		
AGAP002902	2	FBgn0011655	1-to-1	medea	transforming growth factor beta receptor signaling pathway	female ovary	6,8
AGAP002915	2	FBgn0004598	1-to-1	furin 2	proteolysis		6

AGAP002931	2	FBgn0013749	1-to-1	ADP ribosylation factor 102F	protein ADP-ribosylation	+	
AGAP002936	2	FBgn0040900	1-to-many		-		
AGAP002936	2	FBgn0035554	1-to-many		-		
AGAP002942	2	FBgn0015542	1-to-1	similar	regulation of transcription, DNA-dependent		
AGAP002943	2	FBgn0038788	1-to-1	irt2	dendrite morphogenesis		8
AGAP002974	2	FBgn0000036	1-to-1	nicotinic acetylcholine receptor alpha 96Aa	ion transport		6,9
AGAP003018	2	FBgn0003392	1-to-1	shibire	epithelial cell migration, open tracheal system	+	
AGAP003039	2	FBgn0034479	1-to-1	CG8654	transmembrane transport	+	6
AGAP003059	2	FBgn0010113	1-to-1	headcase	negative regulation of terminal cell fate specification, open tracheal system	+	
AGAP003070	2	FBgn0039688	1-to-1	Kuzbanian-like	proteolysis		
AGAP003140	2	FBgn0015269	1-to-1	neurofibromin 1	positive regulation of adenylate cyclase activity	+	
AGAP003162	2	FBgn0085378	1-to-many	CG34349	-		7
AGAP003164	2	FBgn0085378	1-to-many	CG34349	-		7
AGAP003188	2						
AGAP003212	2	FBgn0035529	1-to-1	CG1319	-		
AGAP003271	2	FBgn0037726	1-to-1	CG9492	microtubule-based movement		
AGAP003305	2	FBgn0083949	1-to-1	CG34113	-		6
AGAP003309	2R	FBgn0011281	many-to-many	heromone-binding protein-related protein 3	sensory perception of chemical stim lus	+	
		FBgn0010403	many-to-		sensory		

			many		perception of chemical stimulus		
AGAP003311	2	FBgn0053108	1-to-1	CG33108	-	+	
AGAP003335	2	FBgn0038880	1-to-1	SIFamide receptor	G-protein coupled receptor protein signaling pathway		6
AGAP003509	2	FBgn0041605	1-to-1	complexin	synaptic vesicle exocytosis		
AGAP003572	2	FBgn0259821	1-to-1	cirl	-		
AGAP003584	2R	FBgn0022359	many-to-many	sorbitol dehydrogenase-2	oxidation-reduction process		
		FBgn0024289	many-to-many		oxidation-reduction process		
AGAP003586	2	FBgn0026409	1-to-many	mitochondrial phosphate carrier protein	phosphate ion transport	F	6
AGAP003607	2	FBgn0039294	1-to-1	Cad96Cb	calcium-dependent cell-cell adhesion		6
AGAP003631	2	FBgn0033058	1-to-1	CG14593	G-protein coupled receptor protein signaling pathway		6,9
AGAP003632	2	FBgn0036514	1-to-1	CG12301	mushroom body development	F	
AGAP003658	2	FBgn0028961	1-to-1	allatostatin receptor	G-protein coupled receptor protein signaling pathway		6,9
AGAP003674	2	FBgn0004863	1-to-1	C15	egulation of transcription, DNA-dependent		8,11
AGAP003709	2	FBgn0003429	1-to-1	slowpoke	potassium ion transport		
AGAP003711	2	FBgn0028671	1-to-1	Vha100-1	ATP hydrolysis coupled proton transport	+	
AGAP003725	2	FBgn0039208	1-to-1	Esyt2	-	+	7
AGAP003794	2R	FBgn0004227	1-to-many	no on or off transient A	male courtship behavior, veined wing generated	+	

					song production		
		FBgn0015520	1-to-m ny		nuclear m NA splicing, via spliceosome		
		FBgn0040045	1-to-m ny		NA splicing		
AGAP003856	2	FBgn0038545	1-to-1	CG7713	-	+	
AGAP003857	2	FBgn0004436	1-to-1	ubiquitin conjugating enzyme	entrosome organization	+	
AGAP003871	2	FBgn0025360	1-to-1	optix	ompound eye photoreceptor cell differentiation		8
AGAP003945	2	FBgn0011224	1-to-1	hephaestus	spermatid development		11
AGAP003951	2						
AGAP003958	2						
AGAP003959	2						
AGAP003994	2	FBgn0263289		scri led	M lpighian tubule development		
AGAP003997	2	FBgn0250823	1-to-1	gilgamesh	spermatogenesis	+	
AGAP004000	2	FBgn0000317		rinkled	sensory perception of sound	+	
AGAP004031	2	FBgn0010516	1-to-1	walrus	M lpighian tubule morphogenesis	+	
AGAP004033	2	FBgn0037525	1-to-1	CG17816	-	m le head	
AGAP004046	2	FBgn0037292	1-to-1	CG2022	-	+	6
AGAP004048	2	FBgn0259982	1-to-many	lethal (2) 35Cc	protein folding	F	
AGAP004052	2	FBgn0004595	1-to-1	prospero	7 cell fate commitment	M	8
AGAP004071	2	FBgn0024238	1-to-1	fimbrin	female meiosis chromosome segregation	male testis	
AGAP004074	2						
AGAP004075	2						

AGAP004106	2	FBgn0085406	1-to-1	shal K[+] channel interacting protein	-		
AGAP004160	2	FBgn0085431		CG34402	-		
AGAP004161	2	FBgn0038294		myofilin	skeletal muscle myosin hick filament assembly		
AGAP004163	2	FBgn0038029	1-to-1	CG17639	-		
AGAP004216	2	FBgn0016792	1-to-1		peripheral nervous system development		
AGAP004273	2	FBgn0013334	1-to-1	Synapse-associated protein 47kD	synaptic transmission	+	
AGAP004288	2	FBgn0051108	1-to-1	CG31108	protein modifica ion process	male es is	
AGAP004317	2	FBgn0003450		snake	dorsal/ventral axis specification		
AGAP004341	2	FBgn0027951	1-to-1	TA1-like	phagocytosis, engulfment		
AGAP004405	2	FBgn0011725	1-to-1	twin	nuclear-transcribed mRNA poly(A) tail shortening	+	
AGAP004453	2	FBgn0053517	1-to-1	dopamine 2-like receptor	G-protein coupled receptor protein signaling pathway		6
AGAP004507	2	FBgn0051048	1-to-1	CG31048	-	+	
AGAP004533	2	FBgn0030521	1-to-1	CG10992	autophagic cell death	+	11
AGAP004534	2	FBgn0030521		CG10992	autophagic cell death	F	11
AGAP004592	2	FBgn0004587	1-to-1	B52	nuclear mRNA splicing, via spliceosome	F	
AGAP004593	2	FBgn0016693	1-to-1	putative achaete scute target 1	endocytosis	F	
AGAP004619	2	FBgn0003118	1-to-1	pointed	as protein signal transduction	male es is	

AGAP004640	2R	FBgn0039150	1-to-1	CG13605	-	+	
AGAP004644	2R	FBgn0038447	1-to-1	CG14892	proteolysis		
AGAP004691	2	FBgn0013764	1-to-1	chip	axon guidance	F	8
AGAP004694	2	FBgn0033110		CG9447	-		
AGAP004695	2	FBgn0030894	1-to-1	CG7192	-	F	
AGAP004696	2	FBgn0000611	1-to-1	extradenticle	brain development	F	8
AGAP004701	2	FBgn0040387	1-to-1	Protostome-specific GEF	mushroom body development		7
AGAP004707	2	FBgn0260993	1-to-1	paralytic	response to DDT	F	6
AGAP004716	2						6
AGAP004717	2	FBgn0035192	1-to-1	CG9194	lateral inhibition		
AGAP004718	2	FBgn0033257	1-to-1	CG8713	potassium ion transmembrane transport	+	6
AGAP004721	2	FBgn0260475	1-to-many	CG18278	- acetylglucosamine metabolic process	+	
AGAP004721	2	FBgn0033836	1-to-many		- acetylglucosamine metabolic process		
AGAP004722	2	FBgn0085275	1-to-1	CG34246	-	+	
AGAP004723	2	FBgn0030082		1b	-	F	8
AGAP004724	2						
AGAP004726	2	FBgn0029708	1-to-1	CG3556	cobalamin transport	male es is	
AGAP004727	2	FBgn0041243	1-to-1	gustatory receptor 43a	sensory perception of taste		6
AGAP004728	2	FBgn0052668	1-to-1	CG32668	-	+	
AGAP004731	2	FBgn0029720	1-to-1	CG3009	phospholipid metabolic process		
AGAP004733	2	FBgn0030417	1-to-1	CG15725	-	+	

AGAP004742	2L	FBgn0027580	1-to-1	CG1516	pyruvate metabolic process	F
AGAP004744	2L	FBgn0037643	1-to-1	skpA associated protein	tricarboxylic acid cycle	F
AGAP004745	2L	FBgn0053097	1-to-1	CG42724		
AGAP004750	2L	FBgn0010488	1-to-1	AT1	autophagic cell death	F
AGAP004751	2L	FBgn0051989	1-to-1	chromosome associated protein D3	chromosome organization	F
AGAP004761	2L	FBgn0260963	1-to-1	multiple wing hairs	nt receptor signaling pathway	male Acp
AGAP004762	2L	FBgn0053966	many-to-many	CG12926	-	
		FBgn0053514	many-to-many		neurogenesis	
		FBgn0033437	many-to-many		transport	
		FBgn0053965	many-to-many		-	
		FBgn0051636	many-to-many		transport	
		FBgn0050339	many-to-many		transport	
		FBgn0033434	many-to-many		-	
		FBgn0039107	many-to-many		transport	
		FBgn0039106	many-to-many		-	
AGAP004763	2L	FBgn0053966	many-to-many	CG33965	-	female midgut
		FBgn0053514	many-to-many		neurogenesis	
		FBgn0033437	many-to-many		transport	
		FBgn0053965	many-to-many		-	

		FBgn0051636	many- o-many		ransport		
		FBgn0050339	many- o-many		ransport		
		FBgn0033434	many- o-many		-		
		FBgn0039107	many- o-many		ransport		
		FBgn0039106	many- o-many		-		
AGAP004768	2L	FBgn0033562	1-to-1	CG6751		F	
AGAP004769	2L	FBgn0025806	apparent_1-to-1	Ras-associated protein 2-like	germ-line stem cell maintenance	F	
AGAP004775	2L	FBgn0000455	1-to-1	dipeptidase C	proteolysis	F	
AGAP004776	2L	FBgn0039626	1-to-1	lu7	nuclear mRNA splicing, via spliceosome	F	
AGAP004780	2L	FBgn0037093	1-to-1	CG7597	protein phosphorylation	F	
AGAP004783	2L						6
AGAP004784	2L	FBgn0023550	1-to-many	CG4020	oxidation-reduction process		
AGAP004787	2L	FBgn0023550	1-to-many	CG4020	oxidation-reduction process	+	
AGAP004793	2L	FBgn0022774	1-to-1	ornithine aminotransferase precursor	ornithine metabolic process	+	
AGAP004795	2L	FBgn0014037	1-to-1	u(Tpl)	wing disc dorsal/ventral pat ern formation	F	8
AGAP004800	2L						
AGAP004824	2L	FBgn0026259	1-to-1	eIF5B	ranslational initiation	+	
AGAP004825	2L	FBgn0027500	1-to-1	spindle defective 2	mitotic spindle organization	F	
AGAP004827	2L	FBgn0028496	1-to-1	CG30116	-		
AGAP004833	2L	FBgn0027587	1-to-many	CG7028	protein		

					phosphorylation		
AGAP004845	2L	FBgn0035091	1-to-1	CG3829	defense response	+	6 11
AGAP004846	2L	FBgn0010435	1-to-1	epithelial membrane protein	defense response		6,11
AGAP004847	2L			CG10345			6
AGAP004851	2L	FBgn0035325	many-to-many	CG13806	hitin metabolic process		
		FBgn0052302	many-to-many		hitin metabolic process		
AGAP004855	2L	FBgn0039495	many-to-many	easter	proteolysis	+	
		FBgn0033439	many-to-many		proteolysis		
		FBgn0260474	many-to-many		proteolysis		
		FBgn0039494	many-to-many		proteolysis		
AGAP004856	2L	FBgn0039489	1-to-1	CG5880	protein palmitoylation	F	
AGAP004857	2L	FBgn0003450		snake	spermatogenesis		
AGAP004862	2L	FBgn0035935	1-to-1	misfire	spermatogenesis	male testis	7
AGAP004863	2L	FBgn0035934	1-to-1	transient receptor potential A1	calcium ion transport	m le midgut	6
AGAP004865	2L	FBgn0034046	1-to-1	tungus	learning or memory	male testis	
AGAP004871	2L	FBgn0260657	1-to-1	CG42540	-		6
AGAP004886	2L	FBgn0027066	1-to-1	Eb1	sensory organ development	F	
AGAP004890	2L	FBgn0036030	1-to-1	CG6767	ribonucleoside monophosphate iosynthetic process	F	
AGAP004892	2L	FBgn0016081	1-to-1	furry	rhabdomere development	+	
AGAP004893	2L	FBgn0030562		CG9400	-	+	
AGAP004908	2L	FBgn0025613	1-to-1	CG3081	-		

AGAP004917	2	FBgn0050280		CG30280	signal transduction		
AGAP004918	2	FBgn0050280		CG30280	signal transduction		
AGAP004943	2	FBgn0036391	1-to-1	CG17364	microtubule-based process		
AGAP004961	2	FBgn0086408	1-to-1	stall	ovarian follicle cell stalk form tion	+	
AGAP004963	2	FBgn0000405	1-to-many	cyclin B	cytokinesis	F	
AGAP005031	2	FBgn0003612	1-to-1	suppressor of variegation 2-10	chromosome condensation	F	
AGAP005042	2	FBgn0024187	1-to-1	sunday driver	xon cargo transport		
AGAP005064	2						
AGAP005072	2	FBgn0261451		terribly reduced optic lobes	m inten nce of epithelial cell pical/basal polarity		
AGAP005076	2	FBgn0027356	1-to-1	mphiphysin	endocytosis		
AGAP005091	2	FBgn0042135	1-to-1	CG18812	-		
AGAP005099	2	FBgn0015524	1-to-1	orthopedia	central nervous system development		8
AGAP005123	2	FBgn0030573	1-to-1	nmdyn-D6	CTP biosynthetic process		
AGAP005124	2	FBgn0010548	1-to-1	aldehyde dehydrogenase type III	oxidation-reduction process	+	
AGAP005139	2	FBgn0036273	1-to-1	CG10426	dephosphorylation		
AGAP005164	2	FBgn0034071	1-to-1	CG8405	wing disc dorsal/ventral pattern form tion	+	
AGAP005171	2	FBgn0033243		CG14763	microtubule-based movement		
AGAP005172	2						
AGAP005175	2	FBgn0033246	1-to-1	CG11198	f tty acid biosynthetic		

					process		
AGAP005205	2L	FBgn0035975		peptidoglycan recognition protein LA	detection of bacterium	+	
AGAP005213	2L	FBgn0030252	1-to-1	unconventional myosin class XV	-	+	
AGAP005219	2L	FBgn0261046	1-to-1	Dscam3	cell adhesion		
AGAP005229	2L	FBgn0035329	1-to-many	dromyosuppressin receptor 2	negative regulation of muscle contraction		6
AGAP005229	2L	FBgn0035331	1-to-many		negative regulation of muscle contraction		
AGAP005243	2L	FBgn0085426	1-to-1	gk3	small GTPase mediated signal transduction		
AGAP005244	2L	FBgn0263102		pipsqueak	anterior/posterior axis specification, embryo	male testis	8
AGAP005245	2L	FBgn0005630	1-to-1	longitudinals lacking	axon guidance	F	
AGAP005248	2L	FBgn0050158	1-to-1	CG30158	small GTPase mediated signal transduction		
AGAP005257	2L	FBgn0262868		CG34365	lateral inhibition		
AGAP005285	2L	FBgn0085412		CG34383	-	male testis	
AGAP005287	2L	FBgn0030514	1-to-1	CG9941	-	male testis	
AGAP005288	2L	FBgn0053555	1-to-1	bitesize	positive regulation of multicellular organism growth		
AGAP005292	2L	FBgn0085383	1-to-many	CG34354	-		
AGAP005292	2L	FBgn0085391	1-to-many		regulation of alternative nuclear m A splicing, via spliceosome		
AGAP005300	2L	FBgn0000338	1-to-1	cap-n-collar		male testis	8
AGAP005301	2L	FBgn0039064	1-to-1	CG4467	proteolysis	+	

AGAP005328	2						
AGAP005329	2	FBgn0035171	1-to-1	CG12502	-		
AGAP005397	2						
AGAP005490	2	FBgn0261041	1-to-1	straightjacket	synaptic vesicle fusion to presynaptic membrane		
AGAP005498	2	FBgn0052056		scramblase 1	phospholipid scrambling		
AGAP005507	2	FBgn0013342	1-to-1	n-synaptobrevin	neurotransmitter secretion		
AGAP005508	2	FBgn0053558	1-to-1	CG33558	actin filament organization		6
AGAP005509	2	FBgn0050157	1-to-1	CG30157	-		
AGAP005510	2	FBgn0033463	1-to-1	CG1513	-		
AGAP005546	2	FBgn0052333	1-to-1	CG32333	-	male testis	
AGAP005564	2	FBgn0259211	1-to-1	grainy head	plasma membrane organization		8
AGAP005580	2	FBgn0051008	1-to-many	CG5010	-	male testis	
AGAP005580	2	FBgn0051007	1-to-many		lateral inhibition		
AGAP005581	2	FBgn0034390		CG15093	cellular amino acid metabolic process	F	
AGAP005661	2	FBgn0001078	1-to-1	ftz transcription factor 1	periodic partitioning	+	6,8
AGAP005674	2	FBgn0027111	1-to-1	miple	-		
AGAP005681	2	FBgn0035538	1-to-1	DopEcR	G-protein coupled receptor protein signaling pathway		6
AGAP005719	2	FBgn0010280	1-to-1	TBP-as ociated factor 4	regulation of transcription, DNA-dependent		
AGAP005721	2	FBgn0035287	1-to-1	CG13937	carbohydrate biosynthetic proces	+	6
AGAP005727	2	FBgn0250756	1-to-1	Dbx	-	male testis	8

AGAP005728	2	FBgn0261243		puromycin sensitive aminopeptidase	proteolysis	F	
AGAP005755	2	FBgn0005658	apparent_1-to-1	Ets at 65A	regulation of transcription, DNA-dependent		8,10
AGAP005777	2	FBgn0036566	1-to-1	chloride channel-c	transmembrane transport	+	6
AGAP005778	2	FBgn0035445	1-to-1	CG12014	metabolic process	+	
AGAP007069	2L	FBgn0038366	many-to-many	CG4576	-		
		FBgn0033110	many-to-many		-		
AGAP007088	2	FBgn0034753	1-to-1	CG2852	multicellular organism reproduction	F	
AGAP007124	2	FBgn0022740	1-to-1	54F	visceral muscle development		8
AGAP007163	2	FBgn0034886	1-to-1	phosphodiesterase 8	mesoderm development	+	8
AGAP007207	2	FBgn0003391	1-to-many	shotgun	homophilic cell adhesion		6
AGAP007327	2	FBgn0000567	1-to-1	Ecdysone-induced protein 74EF	salivary gland cell autophagic cell death		8,10
AGAP007329	2						
AGAP007330	2						
AGAP007471	2	FBgn0029114		tollo	defense response		
AGAP007472	2	FBgn0023001	1-to-1	melted	sequestering of triglyceride	male testis	8
AGAP007473	2	FBgn0013269	1-to-1	FK506-binding protein 1	protein folding	F	
AGAP007567	2	FBgn0033476	1-to-1	oysgedart	sperm individualization	male testis	6
AGAP007600	2	FBgn0050421	1-to-1	CG30421	ubiquitin-dependent protein catabolic process		11
AGAP007718	2	FBgn0033502	1-to-1	CG12910	-	+	

AGAP007736	3	FBgn0032694	1-to-1	misexpression suppressor of ras 3	-	M	
AGAP007929	3	FBgn0035537	1-to-1	CG11342	-		
AGAP007952	3	FBgn0015320	1-to-1	ubiquitin conjugating enzyme 2	lipid storage	+	
AGAP008070	3	FBgn0035781		CG8560	proteolysis	male Acp	
AGAP008528	3	FBgn0026713	1-to-1	lethal (1) G0007	nuclear mRNA splicing, via spliceosome	+	
AGAP008534	3	FBgn0027491	1-to-many	Cdk5 activator-like protein	protein phosphorylation	fem le ovary	
AGAP008584	3	FBgn0027611	1-to-1	CG6206	m nnose metabolic process	F	
AGAP008630	3R	FBgn0032668	many-to-many	CG12560	metabolic process	F	
		FBgn0032670	many-to-many		metabolic process		
		FBgn0032669	many-to-many		-		
AGAP008656	3	FBgn0011259	1-to-1	ema-1a	xon midline choice point recognition		
AGAP008681	3						
AGAP008697	3	FBgn0031736	1-to-1	CG11030	-		
AGAP008699	3						
AGAP008712	3	FBgn0051634	1-to-1	organic anion transporting polypeptide 26F	organic anion transport	male testis	
AGAP008713	3						
AGAP008721	3	FBgn0029932	1-to-many	CG8249	transmembrane transport	+	6
AGAP008721	3	FBgn0034045	1-to-many		transmembrane transport		
AGAP008726	3	FBgn0000320	1-to-1	eyes absent	eye-antennal disc morphogenesis		11
AGAP008734	3	FBgn0032233	1-to-1	dpr19	sensory perception of chemical stim lus		

AGAP008741	3R	FBgn0085409	many-to-many	CG34380	signal transduction		
		FBgn0053531	many-to-many		protein phosphorylation		
AGAP008750	3	FBgn0051721	1-to-many	rim9	axon midline choice point recognition	male testis	
AGAP008778	3	FBgn0051814	1-to-1	CG31814	-		
AGAP008851	3	FBgn0086442	1-to-1	mind bomb 2	myoblast fusion		
AGAP008862	3	FBgn0250816	1-to-1	argonaute 3	posttranscriptional gene silencing by A	F	
AGAP008869	3	FBgn0015376	1-to-1	cutlet	cell proliferation	+	
AGAP008871	3	FBgn0027342	1-to-1	frizzled 4	signal transduction		6
AGAP008879	3	FBgn0032262	1-to-many	CG7384	-	F	
AGAP008879	3	FBgn0031201	1-to-many		-		
AGAP008966	3						
AGAP009025	3	FBgn0051612	1-to-1	CG31612	lateral inhibition	+	8
AGAP009038	3	FBgn0028540	1-to-many	CG9008	Golgi organization		
AGAP009052	3R	FBgn0034758	many-to-many	CG13510	-	M	
		FBgn0034759	many-to-many		-		
		FBgn0260767	many-to-many		-		
AGAP009064	3	FBgn0029905	1-to-many	nuclear factor Y-box C	wing disc pattern formation		8
AGAP009067	3	FBgn0004811	1-to-1	female sterile (2) ltoPP43	chorion-containing eggshell formation	F	
AGAP009075	3	FBgn0014189	1-to-1	helicase at 25E	nuclear mRNA splicing, via spliceosome	F	
AGAP009094	3	FBgn0032393	1-to-many	CG12264	iron-sulfur cluster assembly	M	

AGAP009094	3	FBgn0051864	1-to-many		metabolic process		
AGAP009155	3	FBgn0032860	1-to-1	CG15130	-		
AGAP009156	3	FBgn0031285	1-to-1	CG3662	multicellular organism reproduction		
AGAP009207	3	FBgn0003256	1-to-1	p38b	protein phosphorylation	male testis	
AGAP009217	3	FBgn0036891		CG9372	proteolysis		
AGAP009240	3	FBgn0015032		cytochrome P450-4c3	oxidation-reduction process	+	
AGAP009245	3	FBgn0260995	1-to-many	dpr9	-		
AGAP009295	3	FBgn0038296	1-to-1	CG6752	-	+	
AGAP009333	3	FBgn0031359	1-to-1	CG18317	mitochondrial transport	+	6
AGAP009407	3	FBgn0030421		CG3812	phospholipid biosynthetic process		
AGAP009477	3	FBgn0043904	1-to-1	arrest	-		
AGAP009494	3	FBgn0005660	1-to-1	Ets at 21C	dendrite morphogenesis		8,10
AGAP009522	3	FBgn0259735	1-to-1	CG42389	-		
AGAP009531	3	FBgn0032911	1-to-1	torn and diminished rhabdomeres	amino acid transmembrane transport		6
AGAP009533	3	FBgn0085403	1-to-1	Rapgap1	as protein signal transduction		
AGAP009566	3	FBgn0020309	1-to-1	crooked legs	regulation of transcription from A polymerase II promoter		8,11
AGAP009600	3	FBgn0001986	1-to-1	lethal (2) 35Df	nuclear mRNA splicing, via spliceosome	F	
AGAP009627	3						
AGAP009670	3	FBgn0031973	1-to-many	serpin 28D	-		

AGAP009683	3	FBgn0259213		CG42313	-		6
AGAP009777	3	FBgn0025800	1-to-1	smad on X	xon guidance	+	6,8,11
AGAP009799	3	FBgn0028675	1-to-1	sulfonylurea receptor	heart development	M	
AGAP009831	3	FBgn0028527	1-to-1	CG18507	-	M	
AGAP009870	3	FBgn0033729	1-to-1	cuticular protein 49Af	-		
AGAP009876	3	FBgn0050045	1-to-1	cuticular protein 49Aa	-		
AGAP009916	3						
AGAP009917	3					female salivary	
AGAP009918	3						
AGAP009932	3	FBgn0051660	1-to-1	poor gastrulation	metabotropic glutam te receptor signaling pathway		
AGAP009945	3R	FBgn0031418	many-to-many	CG3609	oxidation-reduction process	M	
		FBgn0031417	many-to-many		oxid tion-reduction process		
		FBgn0032609	many-to-many		-		
AGAP010000	3	FBgn0263397		I[[h]] channel	transmembrane transport	M	
AGAP010003	3	FBgn0032252		CG6232	-	F	
AGAP010049	3	FBgn0051665	1-to-1	CG31665	otch signaling pathway		
AGAP010052	3	FBgn0027844	1-to-1	carbonic anhydrase 1	one-carbon metabolic process	M	
AGAP010099	3	FBgn0051876	1-to-many	cuticular protein 30F	neurogenesis		
AGAP010100	3	FBgn0051876	1-to-many	cuticular protein 30F	neurogenesis		
AGAP010110	3	FBgn0031184	1-to-many	CG14615	-	female carcass	
AGAP010111	3	FBgn0031184	1-to-many	CG14615	-	F	
AGAP010115	3	FBgn0259712	1-to-1	CG42366	protein		

					phosphorylation		
AGAP010119	3	FBgn0032125	1-to-many	cuticular protein 30F	-	male malpighian	
AGAP010123	3	FBgn0032125	1-to-many	cuticular protein 30F	-		
AGAP010124	3	FBgn0032125	1-to-many	cuticular protein 30F	-		
AGAP010135	3	FBgn0010583	1-to-1	dreadlocks	axon guidance	F	
AGAP010162	3	FBgn0031518	1-to-1	CG3277	protein phosphorylation		
AGAP010214	3	FBgn0028419	1-to-1	kruppel homolog 2	-	+	6
AGAP010215	3	FBgn0031753	1-to-1	CG13999	-		
AGAP010218	3	FBgn0031752	1-to-1	CG9044	-	+	
AGAP010230	3	FBgn0050021	1-to-1	skiff	regulation of synaptic growth at neuromuscular junction		
AGAP010233	3	FBgn0010395	1-to-1	beta[nu] integrin	cell adhesion	+	6
AGAP010235	3	FBgn0032901	1-to-1	CG9339	regulation of Rab GTPase activity		
AGAP010243	3	FBgn0051954		CG31954	proteolysis	F	
AGAP010244	3	FBgn0085203	1-to-1	CG34174	-	F	
AGAP010245	3	FBgn0002543		leak	axon guidance		
AGAP010254	3	FBgn0033158	1-to-1	CG12164	oxidation-reduction proces		
AGAP010259	3	FBgn0003513	1-to-1	spineles	regulation of transcription, DNA-dependent	male carca	8
AGAP010280	3	FBgn0041713	1-to-1	yellow-c	-		
AGAP010283	3	FBgn0004360	1-to-1	Wnt oncogene analog 2	open tracheal system development		
AGAP010286	3	FBgn0259246		bruchpilot	neurotransmitter secretion		
AGAP010367	3L	FBgn0035375	many-to-many	polypeptide GalNAc transferase 6	oligosaccharide biosynthetic	F	6

					process		
		FBgn0051956	many-to-many		-		
		FBgn0051776	many-to-many		oligosaccharide biosynthetic process		
		FBgn0036529	many-to-many		oligosaccharide biosynthetic process		
		FBgn0036528	many-to-many		oligosaccharide biosynthetic process		
		FBgn0036527	many-to-many		oligosaccharide biosynthetic process		
AGAP010391	3L	FBgn0000242		beadex	imaginal disc-derived wing morphogenesis		
AGAP010394	3L	FBgn0033068	1-to-1	tc-related	-	m le testis	6
AGAP010404	3L	FBgn0010226	1-to-1	glutathione S transferase S1	response to oxidative stress	+	
AGAP010436	3L	FBgn0004852	1-to-1	denylyl cyclase 76E	intracellular signal transduction	+	6
AGAP010452	3L						
AGAP010453	3L	FBgn0033059	1-to-1	CG7845	-	F	
AGAP010454	3L	FBgn0029887	1-to-1	CG3198	nuclear mRNA splicing, via spliceosome	+	
AGAP010473	3L	FBgn0261556		CG42674	cell adhesion	male testis	
AGAP010503	3L	FBgn0029761	1-to-1	small conductance calcium-activated potassium channel	potassium ion transport		6
AGAP010505	3L	FBgn0036080	many-to-many	odorant receptor 67d	G-protein coupled receptor protein signaling pathway		6
		FBgn0037399	many-to-many		sensory perception of smell		

AGAP010508	3L	FBgn0033760	1-to-many	CG8785	mino acid transmem rane transport		
AGAP010512	3L	FBgn0004101	1-to-many	listered	open tracheal system development		8,11
AGAP010513	3L	FBgn0000037	1-to-1	muscarinic acetylcholine receptor 60C	muscarinic cetylcholine receptor signaling pathway		6,9
AGAP010540	3L	FBgn0034804	1-to-1	CG3831	-		
AGAP010596	3L	FBgn0016123	1-to-1	alkaline phosphatase 4	epithelial fluid transport	F	
AGAP010631	3L						
AGAP010637	3L	FBgn0036494		oll-6	defense response	+	
AGAP010638	3L	FBgn0259214	1-to-1	plasma membrane calcium ATPase	cellular calcium ion homeostasis	F	6
AGAP010662	3L	FBgn0038727		CG7432	proteolysis		
AGAP010663	3L	FBgn0038431	1-to-many	CG7829	proteolysis		
AGAP010670	3L	FBgn0027602	1-to-1	CG8611	-	F	
AGAP010743	3L						
AGAP010771	3L						
AGAP010819	3L	FBgn0041182		thiolester containing protein II	ntibacterial humoral response		
AGAP010827	3L	FBgn0033051	many-to-many	death associated molecule related to Mch2	programmed cell death	fem le ovary	11
		FBgn0033659	many-to-many		poptosis		
AGAP010828	3L	FBgn0033051	many-to-many	dream	progr mmed cell death	F	11
		FBgn0033659	many-to-many		poptosis		
AGAP010834	3L	FBgn0003464		small optic lobes	visual behavior		11
AGAP010850	3L	FBgn0085371	1-to-many	CG34342	-		

AGAP010872	3	FBgn0031631	1-to-1	CG3225	nuclear mRNA splicing, via spliceosome		
AGAP010874	3	FBgn0025743	1-to-1	mushroom bodies tiny	mushroom body development		
AGAP010901	3	FBgn0035788		CG8541	-	+	
AGAP010909	3	FBgn0011481	1-to-1	Sequence-specific single-stranded DNA-binding protein	regulation of transcription, DNA-dependent	+	8
AGAP010970	3	FBgn0005631	1-to-many	roundabout	axon guidance		
AGAP010979	3	FBgn0020309		crooked legs	regulation of transcription from A polymerase II promoter	F	
AGAP011008	3	FBgn0037175		CG14455	-		
AGAP011009	3						
AGAP011014	3						
AGAP011015	3	FBgn0037175		CG14455	-		
AGAP011016	3	FBgn0037175		CG14455	-		
AGAP011017	3	FBgn0037175		CG14455	-		
AGAP011227	3	FBgn0085397	1-to-many	Fish-lips	-		
AGAP011293	3	FBgn0033988	1-to-1	parcas	programmed cell death		
AGAP011320	3	FBgn0035010	1-to-1	CG13579	G-protein coupled receptor protein signaling pathway		6
AGAP011323	3	FBgn0052202	1-to-1	CG32202	-	+	
AGAP011360	3	FBgn0040206	1-to-1	kurtz	negative regulation of otch signaling pathway	F	
AGAP011364	3	FBgn0033350	1-to-1	CG8237	-	+	
AGAP011365	3	FBgn0037090	1-to-many	CG9289	-		
AGAP011418	3	FBgn0020309		crooked legs	regulation of transcription from A polymerase II	+	

					promoter		
AGAP011491	3	FBgn0032191	1-to-1	CG5734	cell communication	+	
AGAP011550	3						
AGAP011551	3	FBgn0033339	1-to-1	sec31	RNA splicing	+	
AGAP011618	3	FBgn0037153		olf413	oxidation-reduction proces		
AGAP011628	3						
AGAP011695	3	FBgn0003460	1-to-1	sine oculis	regulation of transcription, DNA-dependent		8
AGAP011705	3	FBgn0025681	1-to-1	CG3558	-		
AGAP011719	3	FBgn0003450		snake	dorsal/ventral axis specification		
AGAP011825	3					+	
AGAP011826	3	FBgn0030610	1-to-1	CG9065	respiratory chain complex IV as embly		
AGAP011846	3	FBgn0032801		CG10165	-		6
AGAP011915	3					female malpighian	6
AGAP011985	3	FBgn0003900	1-to-1	twist	anterior midgut invagination		
AGAP012055	3						
AGAP012102	3						
AGAP012103	3	FBgn0030235	1-to-1	GF-II mRNA-binding protein	-	F	
AGAP012163	3	FBgn0261360	1-to-many	guanylyl cyclase at 76C	-		6
		FBgn0261359	1-to-many		-		
AGAP012189	3	FBgn0034419	1-to-1	CG15111	metabolic process		
AGAP012252	3	FBgn0003091	1-to-many	protein C kinase 53E	protein phosphorylation	F	7
		FBgn0004784	1-to-many		adaptation of rhodopsin		

					mediated signaling		
AGAP012411	3	FBgn0085432	1-to-many	pangolin	embryonic pattern specification		8
AGAP012423	3	FBgn0037098		Wnk	axon guidance	+	
AGAP012424	3	FBgn0013984	1-to-1	Insulin-like receptor	embryonic development via the syncytial blastoderm		6
AGAP013039	2					+	
AGAP013043	X	FBgn0038749	1-to-1	CG4468	-		
AGAP013046	X						
AGAP013055	X	FBgn0260938	1-to-1	tay bridge	adult walking behavior		
AGAP013062	2						
AGAP013073	2	FBgn0028425	1-to-many	hI-21	leucine import	male testis	
AGAP013076	X						
AGAP013140	2						
AGAP013151	2						
AGAP013203	2						
AGAP013238	2	FBgn0051950	1-to-1	CG31950	-		
AGAP013294	2	FBgn0026076	1-to-1	UBL3	-	male testis	
AGAP013327	3	FBgn0038469	1-to-many	peroxidase	oxidation-eduction process	F	
AGAP013412	2	FBgn0003479	1-to-1	spindle A	DNA recombination	F	
AGAP013542	2	FBgn0259722	1-to-1	CG42376	-		
AGAP013553	2						
AGAP013600	3						
AGAP013667	3						
AGAP013686	3						
AGAP013689	2						

AGAP013730	3	FBgn0003310	1-to-1	star	epidermal growth factor receptor signaling pathway	F
AGAP013749	3	FBgn0032125	1-to-many	cuticular protein 30F	-	male m lpighian

[1]Pattern of gene expression from MozAtlas (www.tissue-atl s.org): +, at least one probe detected as expressed; F/M, gender- ia sed expression in fem les (F) or males (M). Cases where expression is unique to one sex and one tissue are explicitly noted.

[2] ee Table 1.

Table S6 ***Anopheles gambiae*** **(A. gam) genes overlapping 1-kb windows with the largest copy number variation (top 0.5% tails) between northern and southern populations, and their** ***D. melanogaster*** **(D. mel) homologs**

Agam Gene ID	Chr	Description (Name)	Dmel FlyBase ID	Homology	Description
AGAP000015	X				
AGAP000017	X		FBgn0086911	1-to-1	Neurobeachin (Protein rugose)(rg)
AGAP000029	X		FBgn0015624	1-to-1	nejire (nej)
AGAP000032	X		FBgn0001250	1-to-1	Integrin alpha-PS2 Precursor (if)
AGAP000035	X				
AGAP000037	X				
AGAP000042	X		FBgn0025639	1-to-1	Histone-lysine N-methyltransferase (Suppressor of variegation 4-20)(Suv4-20)
AGAP000045	X	putative octopamine receptor 1 (G ROAR1)	FBgn0024944	1-to-1	octopamine receptor in mushroom bodies (Oamb)
AGAP000046	X				
AGAP000048	X		FBgn0015589	1-to-many	APC-like (Apc)
			FBgn0026598	1-to-many	adenomatous polyposis coli tumor suppressor homolog 2 (Apc2)
AGAP000053	X		FBgn0037744	1-to-1	CG8417
AGAP000054	X		FBgn0261263	1-to-1	Protein lap4 (Protein scribble)(scrib)
AGAP000057	X		FBgn0005561	1-to-1	shaven (sv)
AGAP000058	X		FBgn0004607	1-to-1	Zinc finger protein 2 (zfn2)
AGAP000064	X		FBgn0025741	1-to-1	plexin A
AGAP000065	X		FBgn0025726	1-to-1	unc-13
AGAP000098	X		FBgn0037679	1-to-1	CG8866
AGAP000106	X		FBgn0031713	1-to-1	Putative ubiquinone biosynthesis monooxygenase COQ6
AGAP000107	X		FBgn0035113	1-to-1	Transient receptor potential channel pyrexia (pyx)
AGAP000121	X		FBgn0038606	1-to-1	CG15803
AGAP000124	X		FBgn0259166	1-to-1	CG42271
AGAP000126	X		FBgn0030808	1-to-1	RhoGAP15B
AGAP000136	X				
AGAP000147	X		FBgn0030864	1-to-1	CG8173
AGAP000169	X		FBgn0038889	1-to-1	CG7922
AGAP000171	X		FBgn0030809	1-to-1	E3 ubiquitin-protein ligase UBR1
AGAP000178	X		FBgn0031266	1-to-1	CG2807
AGAP000179	X		FBgn0004901	1-to-many	Amidophosphoribosyltransferase

				Precursor (Prat)
		FBgn0041194	1-to-many	phosphoribosylamidotransferase 2 (Prat2)
AGAP000182	X			
AGAP000189	X	FBgn0026262	1-to-1	bip2
AGAP000198	X			
AGAP000209	X	FBgn0029830	1-to-1	glutamate receptor binding protein (Grip)
AGAP000222	X	FBgn0053208	1-to-1	molecule interacting with CasL (Mical)
AGAP000233	X	FBgn0030961	1-to-1	CG7058
AGAP000239	X	FBgn0030914	1-to-1	CG6106
AGAP000246	X	FBgn0031196	1-to-1	CG17599
AGAP000252	X	FBgn0039252	1-to-1	CG11771
AGAP000253	X			
AGAP000266	X	FBgn0029749	1-to-1	CG15786
AGAP000267	X			
AGAP000271	X	FBgn0001337	1-to-1	exportin 6 (Exp6)
AGAP000273	X	FBgn0040236	1-to-1	c11.1
AGAP000274	X			
AGAP000286	X			
AGAP000287	X			
AGAP000294	X	FBgn0030156	1-to-1	CG15247
AGAP000300	X	FBgn0003189	1-to-1	CAD protein (Protein rudimentary)
AGAP000301	X	FBgn0015336	1-to-1	CG15865
AGAP000309	X	FBgn0030884	1-to-1	CG6847
AGAP000314	X	FBgn0085430	1-to-1	CG34401
AGAP000326	X	FBgn0030352	1-to-1	UPF0551 protein CG15738 homolog, mitochondrial Precursor
AGAP000331	X			
AGAP000332	X	FBgn0003977	1-to-1	rotein virilizer (vir)
AGAP000337	X	FBgn0029992	1-to-1	Upf2
AGAP000360	X	FBgn0259677	1-to-1	CG42346
AGAP000361	X	FBgn0260005	1-to-1	wtrw
AGAP000363	X	FBgn0016975	1-to-1	stoned B (stnB)
AGAP000364	X	FBgn0016976	1-to-1	Protein stoned-A (Stn-A)
AGAP000365	X	FBgn0035217	apparent_1-to-1	FucTD
AGAP000368	X	FBgn0052743	1-to-1	erine/threonine-protein kinas e (Smg1)
AGAP000374	X	FBgn0004391	1-to-1	shattered (shtd)

AGAP000377	X		FBgn0024997	1-to-1	CG2681
AGAP000378	X		FBgn0030520	1-to-1	CG10990
AGAP000387	X		FBgn0024994	1-to-1	csat (Csat)
AGAP000389	X				
AGAP000390	X		FBgn0030747	1-to-many	CG4301
			FBgn0030746	1-to-many	CG9981
AGAP000397	X		FBgn0010926	1-to-1	Nucleolar protein 14 homolog (l(3)07882)
AGAP000403	X		FBgn0015778	1-to-1	rasputin (rin)
AGAP000406	X		FBgn0026181	1-to-1	Rho-kinase (rok)
AGAP000418	X		FBgn0259680	1-to-1	Putative protein kinase C delta type homolog (Pkcdelta)
AGAP000419	X		FBgn0085446	a parent_1-to-1	CG34417
AGAP000421	X		FBgn0028343	1-to-many	lethal (1) G0222
AGAP000422	X		FBgn0031005	1-to-1	heparan sulfate 3-O sulfotransferase-B (Hs3st-B)
AGAP000427	X		FBgn0033880	1-to-1	CG6553
AGAP000447	X		FBgn0030217	1-to-1	CG2124
AGAP000453	X		FBgn0000392	1-to-1	Oskar ribonucleoprotein complex 147 kDa subunit (cup)
AGAP000461	X		FBgn0259240	1-to-1	tenascin accessory (Ten-a)
AGAP000472	X		FBgn0003463	1-to-1	Dorsal-ventral patterning protein (Sog)
AGAP000473	X		FBgn0032036	1-to-many	CG13384
			FBgn0026150	1-to-many	aminopeptidase P (ApepP)
AGAP000479	X		FBgn0023514	1-to-1	CG14805
AGAP000480	X		FBgn0038609	1-to-1	CG7671
AGAP000483	X		FBgn0038890	1-to-1	CG7956
AGAP000489	X		FBgn0025936	1-to-1	Eph receptor tyrosine kinase (Eph)
AGAP000501	X		FBgn0259704	1-to-1	CG42358
AGAP000513	X		FBgn0030447	1-to-1	Probable alpha-aspartyl dipeptidase
AGAP000520	X				
AGAP000532	X		FBgn0026428	1-to-1	DAC6
AGAP000534	X		FBgn0052495	1-to-many	CG32495
			FBgn0030882	1-to-many	CG32495
AGAP000540	X		FBgn0030816	1-to-many	CG16700
			FBgn0030817	1-to-many	CG4991
AGAP000545	X		FBgn0040234	1-to-1	c12.2
AGAP000555	X	olybdenum cofactor	FBgn0002641	1-to-1	Molybdenum cofactor sulfurase

		sulfurase (mal)			(M S)(MoCo sulfurase)(Protein maroon-like)(Mal)
AGAP000556	X		FBgn0040650	1-to-1	CG15456
AGAP000560	X		FBgn0024992	1-to-1	CG2658
AGAP000571	X	Clip-Domain Serine Protease, family C (CLIPC5)	FBgn0030926	many-to-many	Serine protease persephone Precursor (psh)
			FBgn0030925	many-to-many	CG6361
AGAP000575	X		FBgn0011606	1-to-many	inesin-like protein at 3A (Klp3A)
AGAP000576	X		FBgn0022153	1-to-1	lethal (2) k05819 (l(2)k05819)
AGAP000579	X		FBgn0031140	1-to-1	PC1b
AGAP000602	X		FBgn0260935	1-to-1	immune response deficient 1 (ird1)
AGAP000615	X		FBgn0028475	1-to-1	CG10221
AGAP000621	X		FBgn0029807	1-to-many	CG3108
			FBgn0029804	1-to-many	CG3097
AGAP000623	X		FBgn0030506	1-to-1	igase4 (Lig4)
AGAP000631	X		FBgn0261383	1-to-1	CG3125
AGAP000656	X		FBgn0030930	1-to-1	N-acetylgalactosaminyltransferase 7 (GalNAc-T2)
AGAP000668	X		FBgn0027620	1-to-many	ATP-dependent chromatin assembly factor large subunit (Acf1)
AGAP000674	X		FBgn0000233	1-to-1	Transcription factor btd (btd)
AGAP000681	X		FBgn0037435	1-to-1	CG18048
AGAP000686	X		FBgn0260933	1-to-1	reduced mechanoreceptor potential A (rempA)
AGAP000708	X		FBgn0030087	1-to-1	Probable phosphorylase b kinase regulatory subunit alpha
AGAP000717	X		FBgn0031011	1-to-1	CG8034
AGAP000740	X		FBgn0035999	1-to-1	CG3552
AGAP000745	X				
AGAP000747	X		FBgn0014006	1-to-1	protein kinase at 92B
AGAP000754	X				
AGAP000788	X		FBgn0004511	1-to-1	dusky (dy)
AGAP000810	X		FBgn0038816	1-to-1	Leucine-rich repeat kinase (Lrrk)
AGAP000869	X		FBgn0000163	1-to-1	bazooka (baz)
AGAP000898	X		FBgn0029157	1-to-1	Protein phosphatase Slingshot (ssh)
AGAP000901	X		FBgn0030478	1-to-1	CG1640
AGAP000903	X		FBgn0000524	1-to-1	Protein deltex (dx)
AGAP000904	X				

AGAP000906	X				
AGAP000930	X		FBgn0029903	1-to-1	pod1
AGAP000973	X		FBgn0039774	1-to-many	eutral ceramidase acylsphingosine deacylase)(Slug-a-bed protein)
AGAP000984	X		FBgn0003053	1-to-1	pebbled
AGAP000990	X		FBgn0035799	1-to-many	CG14838
AGAP000996	X		FBgn0015926	1-to-1	discontinuous actin hexagon (dah)
AGAP000998	X		FBgn0051072	1-to-1	lysosomal enzyme receptor protein (Lerp)
AGAP001004	X	Toll-like Receptor (TOLL1A)	FBgn0003717	1-to-many	rotein toll Precursor (Tl)
AGAP001015	X		FBgn0004647	1-to-1	Neurogenic locus Notch protein (N)
AGAP001018	X		FBgn0036219	1-to-1	CG14127
AGAP001021	X		FBgn0086450	1-to-1	suppressor of rudimentary (su(r))
AGAP001035	X		FBgn0031006	1-to-1	rapamycin-insensitive companion of Tor (rictor)
AGAP001187	2R		FBgn0004198	1-to-1	Homeobox protein cut (ct)
AGAP001189	2R	odorant binding protein (OBP10)	FBgn0034468	many-to-many	General odorant-binding protein 56a (Obp56a)
			FBgn0034470	many-to-many	General odorant-binding protein 56d (Obp56d)
			FBgn0034471	many-to-many	dorant-binding protein 56e (Obp56e)
AGAP001272	2R		FBgn0038769	1-to-1	CG10889
AGAP001296	2R		FBgn0260794	1-to-1	CG42574
AGAP001298	2R		FBgn0039229	1-to-1	CG6995
AGAP001368	2R				
AGAP001386	2R		FBgn0001253	1-to-many	Ecdysone-inducible gene E1 (ImpE1)
			FBgn0026427	1-to-many	Su(var)2-HP2
AGAP001412	2R		FBgn0004644	1-to-1	Protein hedgehog Precursor (Hh)
AGAP001493	X		FBgn0003129	1-to-1	Paired box pox-meso protein (Poxm)
AGAP001535	2R		FBgn0005386	1-to-1	absent (ash1)
AGAP001544	2R				
AGAP001568	2R		FBgn0036242	1-to-many	CG6793
			FBgn0029501	1-to-many	Caldesmon-related protein (Crtp)
AGAP001606	2R		FBgn0034733	1-to-1	CG4752
AGAP001633	2R		FBgn0005666	1-to-1	bent (bt)
AGAP001700	2R				
AGAP001710	2R				
AGAP001759	2R		FBgn0037357	1-to-1	sec23

AGAP001775	2R			
AGAP001780	2R	FBgn0002431	1-to-1	E3 ubiquitin-protein ligase hyd (Protein hyperplastic discs) (hyd)
AGAP001783	2R	FBgn0040208	1-to-1	katanin 60
AGAP001820	2R	FBgn0022787	1-to-1	helicase 89B (Hel89B)
AGAP001888	2R	FBgn0030327	1-to-1	Alpha-(1,6)-fucosyltransferase (FucT6)
AGAP001916	2R	FBgn0053556	1-to-1	formin 3 (form3)
AGAP001945	2R	FBgn0037265	1-to-1	CG12001
AGAP002013	2R	FBgn0035577	1-to-1	CG13708
AGAP002030	2R	FBgn0011785	1-to-1	BRWD3
AGAP002036	2R			
AGAP002038	2R	FBgn0027794	1-to-1	CG14786
AGAP002091	2R	FBgn0000052	1-to-1	Phosphoribosylformylglycinamidine synthase)(Protein adenosine-2) (ade2)
AGAP002110	2R	FBgn0028360	1-to-many	lethal (1) G0148 (l(1)G0148)
		FBgn0032677	1-to-many	CG5790
AGAP002181	2R	FBgn0035948	1-to-1	CG5644
AGAP002215	2R	FBgn0037541	1-to-1	CG2747
AGAP002224	2R	FBgn0017561	1-to-1	Open rectifier potassium channel protein 1 (Two pore domain potassium channel Ork1)
AGAP002233	2R			
AGAP002243	2R	FBgn0039955	1-to-1	CG41099
AGAP002265	2R	FBgn0024285	1-to-1	Srp54
AGAP002268	2R			
AGAP002272	2R	FBgn0085445	1-to-1	Ank2
AGAP002274	2R			
AGAP002288	2R	FBgn0001205	1-to-1	3-hydroxy-3-methylglutaryl-coenzyme A reductase (Hmgcr)
AGAP002293	2R			
AGAP002299	2R	FBgn0034279	1-to-1	CG18635
AGAP002300	2R	FBgn0038269	1-to-1	Rrp6
AGAP002315	2R	FBgn0039705	1-to-1	CG31033
AGAP002336	2R	FBgn0031016	1-to-many	kekkon5 (kek5)
		FBgn0032484	1-to-many	kekkon4 (kek4)
AGAP002351	2R	FBgn0039120	1-to-1	up98
AGAP002469	2R	FBgn0015772	1-to-1	Numb-associated kinase (Nak)
AGAP002502	2R	FBgn0260634	1-to-many	eukaryotic translation initiation factor 4G2 (eIF4G2)

			FBgn0023213	1-to-many	eukaryotic translation initiation factor 4G (eIF4G)
AGAP002503	2R				
AGAP002579	2R		FBgn0002306	1-to-1	Putative epidermal cell surface receptor Precursor (Stranded at second protein) (sas)
AGAP002585	2R				
AGAP002610	2R		FBgn0029939	1-to-1	CG9650
AGAP002624	2R		FBgn0038312	1-to-1	CG4334
AGAP002638	2R				
AGAP002648	2R		FBgn0035898	1-to-1	CG6915
AGAP002715	2R		FBgn0010110	1-to-1	enhanced adult sensory threshold (east)
AGAP002725	2R		FBgn0030065	1-to-1	CG12075
AGAP002735	2R		FBgn0034975	1-to-1	enoki mushroom (enok)
AGAP002737	2R		FBgn0031273	apparent_1-to-1	CG2839
AGAP002739	2R		FBgn0031879	1-to-1	SP1070
AGAP002741	2R		FBgn0003862	1-to-1	Histone-lysine N-methyltransferase trithorax (trx)
AGAP002760	2R		FBgn0002878	1-to-1	mutagen-sensitive 101 (mus101)
AGAP002877	2R	Tetratricopeptide repeat protein 30	FBgn0032470	1-to-1	Tetratricopeptide repeat protein 30 homolog
AGAP002878	2R				
AGAP002881	2R	putative neuropeptide receptor 1 (GPRNPR1)	FBgn0038874	1-to-1	ETHR
AGAP002891	2R	putative metabotropic glutamate receptor 4 (GPRMG 4)	FBgn0051116	1-to-1	Chloride channel protein 2 (ClC-2) [Source:UniProtKB/Swiss-Prot;Acc:Q9VGH7]
AGAP002915	2R		FBgn0004598	1-to-1	Furin-like protease 2 Precursor (Fur2)
AGAP002951	2R		FBgn0030701	1-to-1	CG16952
AGAP003145	2R		FBgn0036725	1-to-1	
AGAP003151	2R				
AGAP003174	2R		FBgn0037804	1-to-1	CG11870
AGAP003184	2R		FBgn0020764	1-to-1	aminolevulinate synthase (Alas)
AGAP003189	2R		FBgn0052121	1-to-1	CG32121
AGAP003301	2R		FBgn0003525	1-to-many	-phase inducer phosphatase (Cdc25-like protein)(Protein string) (stg)
			FBgn0002673	1-to-many	Cdc25-like protein phosphatase twine

				(twe)
AGAP003320	2	FBgn0037800	1-to-1	CG3996
AGAP003366	2	FBgn0025802	1-to-1	SET domain binding factor (Sbf)
AGAP003453	2	FBgn0039151	1-to-1	CG13607
AGAP003489	2	FBgn0039727	1-to-1	CG15523
AGAP003506	2	FBgn0039528	1-to-1	distracted (dsd)
AGAP003568	2	FBgn0083963	1-to-1	CG34127
AGAP003610	2			
AGAP003663	2	FBgn0003261	1-to-1	ATP-dependent RNA helicase p62 (Rm62)
AGAP003827	2	FBgn0051151	1-to-1	otein winged eye (wge)
AGAP003849	2	FBgn0250829	1-to-1	Poly-glutamine tract binding protein 1 (PQBP-1)
AGAP003918	2	FBgn0039907	1-to-1	otein BCL9 homolog (Protein legless) (lgs)
AGAP005175	2L	FBgn0033246	1-to-1	CG11198 [
AGAP005213	2L	FBgn0030252	1-to-1	unconventional myosin class XV (Myo10A)
AGAP005273	2L	FBgn0039728	1-to-1	CG7896
AGAP005361	2L			
AGAP005362	2L	FBgn0035518	1-to-1	CG15011
AGAP005391	2L	FBgn0011802	1-to-1	Gemin3 (Gem3)
AGAP005396	2L			
AGAP005472	2L	FBgn0035420	1-to-1	CG14967
AGAP005525	2L			
AGAP005526	2L	FBgn0052702	1-to-1	CG32702
AGAP005536	2L			
AGAP005557	2L	FBgn0035649	1-to-1	CG10483]
AGAP005558	2L	FBgn0038271	1-to-many	CG3731
AGAP005630	2L	FBgn0261014	1-to-1	Transitional endoplasmic reticulum ATPase (TER94)
AGAP005640	2L	FBgn0036828	1-to-1	CG6841
AGAP005779	2L			
AGAP005788	2L			
AGAP005789	2L	FBgn0038953	1-to-many	CG18596
AGAP005795	2L	FBgn0035293	1-to-1	CG5687
AGAP005796	2L	FBgn0003041	1-to-1	pebble (pbl)
AGAP005803	2L	FBgn0259163	1-to-1	CG42268
AGAP005807	2L	FBgn0053205	1-to-many	CG33205

			FBgn0052050	1-to-many	CG32050
AGAP005808	2				
AGAP005816	2		FBgn0035756	1-to-1	unc-13-4A
AGAP005837	2		FBgn0010052	many-to-many	juvenile hormone esterase (jhe)
			FBgn0034076	many-to-many	juvenile hormone esterase duplication (Jhedup)
AGAP005850	2		FBgn0029667	1-to-1	Growth arrest-specific protein 8 homolog (Gas8)
AGAP005895	2		FBgn0035956	many-to-many	Dorsocros 2 (Doc2)
			FBgn0028789	many-to-many	Dorsocros 1 (Doc1)
			FBgn0035954	many-to-many	Dorsocros 3 (Doc3)
AGAP005898	2		FBgn0011817	1-to-1	nemo
AGAP005904	2		FBgn0001104	1-to-1	Guanine nucleotide-binding protein (G-ialpha65A)
AGAP005962	2		FBgn0052372	1-to-1	CG32372
AGAP005963	2		FBgn0035793	1-to-1	CG7546
AGAP005964	2		FBgn0052423	1-to-1	Protein alan shepard (shep)
AGAP006018	2		FBgn0004870	1-to-1	Protein bric-a-brac 1 (bab1)
AGAP006022	2		FBgn0030627	1-to-many	germ cell-expressed bHLH-PAS (gce)
			FBgn0002723	1-to-many	ethoprene-tolerant (Met)
AGAP006025	2		FBgn0035711	1-to-1	CG8519
AGAP006026	2				
AGAP006041	2		FBgn0030693	1-to-many	CG8974
			FBgn0052581	1-to-many	CG32581
			FBgn0052847	1-to-many	CG32847
AGAP006045	2		FBgn0034970	1-to-1	rotein yorkie (yki)
AGAP006082	2		FBgn0033753	1-to-1	Probable cytochrome P450 301a1, mitochondrial Precursor(Cyp301a1)
AGAP006089	2		FBgn0052062	1-to-1	CG32062
AGAP006091	2		FBgn0052066	1-to-1	CG32066
AGAP006115	2		FBgn0036043	1-to-many	CG8177
AGAP006156	2	putative metabotropic glutamate receptor 1 (GPRMG 1)	FBgn0085401	1-to-1	CG34372
AGAP006158	2		FBgn0035073	1-to-1	CG16896

AGAP006186	2L	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum (Ca-P60A)	FBgn0004551	1-to-1	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum ype (Ca-P60A)
AGAP006188	2L		FBgn0035802	1-to-1	CG33275
AGAP006217	2L				
AGAP006218	2L	putative methuselah receptor 4 (GPRMTH4)	FBgn0035132	many-to-many	Probable G-protein coupled receptor h-like 10 Precursor (Protein methuselah-like 10) (mthl10)
			FBgn0023000	many-to-many	G-protein coupled receptor M h Precursor (Protein methuselah) (mth)
			FBgn0028956	many-to-many	Probable G-protein coupled receptor h-like 3 (mthl3)
			FBgn0035623	many-to-many	Probable G-protein coupled receptor h-like 2 (mthl2)
			FBgn0034219	many-to-many	Probable G-protein coupled receptor h-like 4 (mthl4)
			FBgn0035789	many-to-many	Probable G-protein coupled receptor h-like 6 (mthl6)]
			FBgn0035847	many-to-many	robable G-protein coupled receptor h-like 7 (mthl7)
			FBgn0045443	many-to-many	Probable G-protein coupled receptor h-like 11 (mthl11)
			FBgn0045442	many-to-many	Probable G-protein coupled receptor h-like 12 (mthl12)
			FBgn0050018	many-to-many	methuselah-like 13 (mthl13)
			FBgn0035131	many-to-many	Probable G-protein coupled receptor h-like 9 (mthl9)
			FBgn0052475	many-to-many	Probable G-protein coupled receptor h-like 8 (mthl8)
AGAP006270	2L		FBgn0004603	1-to-1	Tyrosine-protein kinase Src42A (Src42A)
AGAP006271	2L		FBgn0033049	1-to-1	CG14471
AGAP006274	2L		FBgn0085414	1-to-1	dpr12
AGAP006275	2L				
AGAP006277	2L				
AGAP006282	2L		FBgn0259878	1-to-1	follistatin (Fs)
AGAP006322	2L	NA-Leu			
AGAP006330	2L		FBgn0042185	1-to-1	CG18769

AGAP006345	2				
AGAP006360	2		FBgn0036663	1-to-1	CG9674
AGAP006404	2				
AGAP006405	2		FBgn0033791	1-to-1	derailed 2 (Drl-2)
AGAP006406	2		FBgn0037626	1-to-many	CG8236
AGAP006436	2		FBgn0003415	1-to-1	Mediator of RNA polymerase II transcription subunit 13 (skd)
AGAP006439	2		FBgn0011591	1-to-1	Fringe glycosyltransferase (fng)
AGAP006440	2				
AGAP006571	2		FBgn0004865	1-to-1	Ecdysone-induced protein 78C (Eip78C)
AGAP006590	2		FBgn0085447	apparent_1-to-1	Protein still life, isoform SIF type 1 (sif)
AGAP006592	2		FBgn0035625	1-to-1	Blimp-1
AGAP006606	2				
AGAP006633	2		FBgn0034135	1-to-1	Syntrophin-like 2 (Syn2)
AGAP006656	2		FBgn0260943	1-to-1	RNA-binding protein Musashi homolog (Rbp6)
AGAP006664	2				
AGAP006665	2		FBgn0005536	1-to-1	myosin binding subunit (Mbs)
AGAP006723	2		FBgn0000326	1-to-many	cricklet (clt)
AGAP006731	2		FBgn0003391	1-to-many	DE-cadherin Precursor (Protein shotgun) (shg)
AGAP006737	2				
AGAP006741	2		FBgn0259224	apparent_1-to-1	CG42324
AGAP006745	2		FBgn0036101	1-to-many	ninjurin A (NijA
AGAP006754	2				
AGAP006763	2		FBgn0033000	1-to-1	CG14464
AGAP006764	2				
AGAP006765	2		FBgn0085512	1-to-1	CG40733
AGAP006776	2				
AGAP006800	2		FBgn0039350	1-to-1	jing interacting gene regulatory 1 (jigr1)
AGAP006839	2	cuticular protein 67 (CPR67)	FBgn0036879	many-to-many	cuticular protein 76Bb (Cpr76Bb)
			FBgn0036878	many-to-many	cuticular protein 76Ba (Cpr76Ba)
AGAP006872	2		FBgn0045980	1-to-1	nimA-like kinase (niki)
AGAP006890	2		FBgn0013750	1-to-1	ADP-ribosylation factor 3 (Arf51F)
AGAP006896	2		FBgn0035768	1-to-1	CG14834

AGAP006914	2				
AGAP006920	2				
AGAP006967	2		FBgn0052521	1-to-1	CG32521
AGAP006995	2		FBgn0030300	1-to-many	sphingosine kinase 1 (Sk1)
			FBgn0052484	1-to-many	sphingosine kinase 2 (Sk2)
AGAP007005	2		FBgn0050089	1-to-1	CG30089
AGAP007006	2		FBgn0028371	1-to-1	jitterbug (jbug)
AGAP007007	2				
AGAP007008	2		FBgn0259145	1-to-1	CG42260
AGAP007017	2		FBgn0036749	many-to-many	CG7460
			FBgn0033584	many-to-many	CG7737
			FBgn0035943	many-to-many	CG5653
			FBgn0036750	many-to-many	CG6034
AGAP007023	2	rotein cueball (cue)			
AGAP007031	2		FBgn0003326	1-to-1	rotein scabrous (sca)
AGAP007049	2		FBgn0034638	1-to-1	CG10433
AGAP007051	2		FBgn0020521	1-to-1	piopio (pio)
AGAP007058	2		FBgn0000157	1-to-1	Homeotic protein distal-less (Dll)
AGAP007059	2				
AGAP007066	2		FBgn0034360	1-to-1	CG10927
AGAP007067	2	tRNA-Met			
AGAP007181	2		FBgn0035347	1-to-1	CG33232
AGAP007185	2				
AGAP007341	2				
AGAP007474	2		FBgn0260442	1-to-1	rhea
AGAP007532	2		FBgn0004397	1-to-1	Vinculin (Vinc)
AGAP007562	2		FBgn0052311	1-to-1	zormin
AGAP007776	3R				
AGAP007801	3R		FBgn0016076	1-to-1	vrille (vri)
AGAP007803	3R		FBgn0051992	1-to-many	gawky (gw)
AGAP007905	3R		FBgn0000497	1-to-1	rotein dachsous (ds)
AGAP007924	3R		FBgn0001075	1-to-1	Cadherin-related tumor sup ressor (ft)
AGAP008178	3R				
AGAP008179	3R	Class B Scavenger Receptor (SCRBQ3)	FBgn0015924	many-to-many	croquemort (crq)

		FBgn0002939	many-to-many	neither inactivation nor afterpotential D (ninaD)
		FBgn0051741	many-to-many	CG31741
AGAP008268	3R	FBgn0003891	1-to-1	Maternal protein tudor (tud)
AGAP008353	3R	FBgn0011232	1-to-1	Vacuolar protein sorting-associated protein 54 (Protein scattered) (scat)
AGAP008362	3R	FBgn0001301	1-to-1	kelch (kel)
AGAP008636	3R	FBgn0010309	1-to-1	rotein pigeon
AGAP008698	3R	FBgn0250786	1-to-1	Chromodomain-helicase-DNA-binding protein 1 (Chd1)
AGAP008752	3R			
AGAP008801	3R	FBgn0022201	1-to-1	Trafficking kinesin-binding protein milt (Protein milton) (milt)
AGAP008826	3R	FBgn0032838	1-to-1	CG13966
AGAP008851	3R	FBgn0086442	1-to-1	mind bomb 2 (mib2)
AGAP008858	3R	FBgn0032348	1-to-1	N domain-containing protein CG4751
AGAP008859	3R	FBgn0024248	1-to-1	Insulin receptor substrate 1 (chico)
AGAP008970	3R	FBgn0010660	1-to-1	uclear pore complex protein up214 (Nup214)
AGAP009028	3R	FBgn0028979	1-to-many	rotein tiptop (tio)
		FBgn0003866	1-to-many	rotein teashirt (tsh)
AGAP009117	3R	FBgn0031893	1-to-1	Calcium uptake protein 1 homolog, mitochondrial Precursor
AGAP009119	3R	FBgn0260484	1-to-many	sc70-interacting protein 2 (HIP-replacement)
		FBgn0029676	1-to-many	sc70-interacting protein 2 (HIP-replacement)
AGAP009176	3R	FBgn0027571	1-to-1	CG3523
AGAP009200	3R	FBgn0000299	1-to-1	Collagen alpha-1(IV) chain (Cg25C)
AGAP009201	3R			
AGAP009332	3R	FBgn0031927	1-to-1	CG13792
AGAP009344	3R			
AGAP009350	3R	FBgn0000097	1-to-1	Ets DNA-binding protein pokkuri (Protein anterior open)(aop)
AGAP009447	3R	FBgn0005278	1-to-1	S-adenosylmethionine synthase (M(2)21AB)
AGAP009522	3R	FBgn0259735	1-to-1	CG42389

AGAP009570	3	FBgn0037707	1-to-many	A-binding protein S1 (RnpS1)
		FBgn0085363	1-to-many	CG34334
AGAP009571	3	FBgn0261532	1-to-1	cadmus (cdm)
AGAP009847	3			
AGAP010021	3			
AGAP010630	3	FBgn0005614	1-to-1	Transient-receptor-potential-like protein (tpl)
AGAP010697	3	FBgn0037344	1-to-many	CG2926]
AGAP010800	3	FBgn0029941	1-to-many	CG1677
		FBgn0051601	1-to-many	CG31601
AGAP010981	3			
AGAP011030	3	FBgn0036511	1-to-1	CG6498
AGAP011138	3			
AGAP011139	3	FBgn0025740	1-to-1	lexin-B (plexB)
AGAP011162	3	FBgn0027948	1-to-1	mini spindles (msps)
AGAP011192	3	FBgn0023518	1-to-1	Histone-lysine N-methyltransferase trr (trr)
AGAP011206	3	FBgn0002036	apparent_1-to-1	rotein anon-37Cs
AGAP011207	3	FBgn0036749	many-to-many	CG7460
		FBgn0033584	many-to-many	CG7737
		FBgn0035943	many-to-many	CG5653
		FBgn0036750	many-to-many	CG6034
AGAP011333	3	FBgn0024806	1-to-1	Disco-interacting protein 2 (DI 2)
AGAP011355	3	FBgn0036454	1-to-1	CG17839
AGAP011378	3	FBgn0010452	many-to-many	tartan (trn)
		FBgn0023095	many-to-many	capricious (caps)
AGAP011396	3	FBgn0013733	1-to-many	hort stop (shot)
AGAP011399	3	FBgn0045862	1-to-1	barentsz (btz)
AGAP011400	3			
AGAP011421	3	FBgn0026179	1-to-1	schizo (siz)
AGAP011454	3	FBgn0033718	1-to-1	anastral spindle 3 (ana3)
AGAP011495	3			

AGAP011545	3		FBgn0036451	1-to-1	CG9425
AGAP011562	3		FBgn0035264	1-to-1	Oseg4
AGAP011640	3				
AGAP011728	3		FBgn0053291	1-to-1	CG33291
AGAP011731	3				
AGAP011764	3		FBgn0034184	1-to-1	CG9646
AGAP011823	3		FBgn0031107	1-to-1	Probable E3 ubiquitin-protein ligase ERC2
AGAP011851	3	Protein KIAA0664 homolog	FBgn0034087	1-to-1	Protein KIAA0664 homolog
AGAP011902	3		FBgn0003254	1-to-1	ribbon (rib)
AGAP011965	3		FBgn0041161	1-to-1	bluestreak (blue)
AGAP012009	3		FBgn0013591	1-to-many	Chromodomain-helicase-DNA-binding protein Mi-2
			FBgn0023395	1-to-many	Chromodomain-helicase-DNA-binding protein 3 (Chd3)
AGAP012013	3		FBgn0044324	1-to-1	chromator (Chro)
AGAP012023	3		FBgn0000307	1-to-1	Protein chiffon (chif)
AGAP012024	3				
AGAP012045	3		FBgn0029979	1-to-1	CG10777
AGAP012288	3		FBgn0015838	1-to-1	Van gogh (Vang)
AGAP012313	3				
AGAP012326	3	Toll-like Receptor (TOLL7)	FBgn0004364	1-to-many	18 wheeler (18w)
			FBgn0034476	1-to-many	Toll-7
AGAP012332	3				
AGAP013009	X				
AGAP013042	X				
AGAP013057	2R				
AGAP013190	2R		FBgn0041723	1-to-1	rhomboid-5 (rho-5)
AGAP013263	2R		FBgn0037304	1-to-1	CG1113
AGAP013357	2R		FBgn0037944	1-to-1	CG6923
AGAP013392	X		FBgn0030055	1-to-1	CG12772
AGAP013411	2R		FBgn0028386	1-to-many	Putative transcription factor capicua (cic)
AGAP013457	2R				
AGAP013458	2R				
AGAP013468	2R				
AGAP013477	2R				
AGAP013495	X		FBgn0030974	1-to-1	CG7358
AGAP013497	X		FBgn0259705	1-to-many	CG42359

AGAP013538	2R	FBgn0043900	1-to-1	Protein pygopus (pygo)

Table S7 Functional annotation clusters of *An. gambiae* genes with the largest copy number variation between northern and southern populations

Annotation cluster	Representative annotation term(s)	Gene count	Enrichment score
1	Imaginal disc-derived appendage morphogenesis	20	2.90
2	Ankyrin repeat	10	2.76
3	ATP-binding, serine/threonine protein kinase activity	60	2.48
4	Compound eye development, regulation of nervous system development, establishment or maintenance of cell or tissue polarity	17	2.08
5	Transmembrane	46	1.83
6	Limb development, im ginal disc pattern form tion, notch sign ling pathw y	10	1.83
7	Transmembrane receptor protein tyrosine kinase signaling pathway	7	1.58
8	EGF-like dom in	14	1.54
9	eucine-rich repeat	9	1.53
10	eproductive developmental process	14	1.51
11	lexin	3	1.49
12	Bromodomain	3	1.33