

A Single Gene Causes Both Male Sterility and Segregation Distortion in *Drosophila* Hybrids

Nitin Phadnis* and H. Allen Orr

A central goal of evolutionary biology is to identify the genes and evolutionary forces that cause speciation, the emergence of reproductive isolation between populations. Despite the identification of several genes that cause hybrid sterility or inviability—many of which have evolved rapidly under positive Darwinian selection—little is known about the ecological or genomic forces that drive the evolution of postzygotic isolation. Here, we show that the same gene, *Overdrive*, causes both male sterility and segregation distortion in F₁ hybrids between the Bogota and U.S. subspecies of *Drosophila pseudoobscura*. This segregation distorter gene is essential for hybrid sterility, a strong reproductive barrier between these young taxa. Our results suggest that genetic conflict may be an important evolutionary force in speciation.

Segregation distorters (also called transmission ratio distorters) are selfish genetic elements that manipulate Mendelian transmission to their own advantage. Because they are inherited by more than 50% of functional gametes, such distorters can spread rapidly through populations. Natural selection, however, favors the suppression of distortion, especially if the distorter locus resides on a sex chromosome and biases sex ratios. Theory predicts that mutations suppressing segregation distortion will invade populations once distorter alleles are at a high frequency or are fixed (1, 2). Bouts of distortion and suppression could occur repeatedly and often. If suppressors of distortion are less than fully dominant, segregation distortion may reappear in F₁ hybrids between species. And because segregation distorters often act by destroying gametes (3), such reexpression could result in hybrid sterility (4, 5). The idea that intragenomic conflict involving segregation distorters may cause hybrid sterility is controversial (4–9). Although it can explain both the observed rapid evolution of most hybrid incompatibility genes (10–15) and the preferential sterility of the heterogametic sex [an aspect of Haldane's rule (16)], direct empirical evidence is scarce [(17) but see (8)].

Drosophila pseudoobscura pseudoobscura and *Drosophila pseudoobscura bogotana* (hereafter USA and Bogota, respectively) are allopatric subspecies that diverged just 155,000 to 230,000 years ago (18). Crosses between Bogota females and USA males produce F₁ hybrid males that are nearly completely sterile, whereas all other F₁ hybrids are fertile (19). Hybrid sterility appears to involve a single complex genetic incompatibility, in which several loci are essential for the expression of full sterility. Bogota alleles at loci on the right and left arms of the X chromosome (XR and XL, respectively) interact with dominant

USA alleles on the second and third autosomes to cause hybrid sterility (20). Because these genes are essential for hybrid sterility, they could not have evolved after the attainment of complete reproductive isolation. These genes may therefore be important components of ongoing speciation.

When aged, F₁ hybrid males (previously thought to be completely sterile) become weakly fertile. These hybrid males produce progeny that are almost all daughters (21). This sex-ratio distortion is not caused by hybrid inviability but by an overrepresentation of X-bearing sperm among functional gametes; consequently, X-bearing sperm from F₁ hybrid males fertilize more eggs than do Y-bearing sperm (21). The precise mechanism of

segregation distortion may involve true meiotic drive (which acts during meiosis) or gamete killing or inactivation (which acts after meiosis); in either case, the X chromosome carrying this segregation distorter enjoys a selective advantage.

Male sterility and segregation distortion in Bogota-USA hybrids both map to the same chromosomal regions, and these regions show a similar pattern of complex epistasis for both phenotypes (20, 21). A region on the Bogota XR that is tightly linked to the visible mutation *sepia* (*se*, XR-156.6) was identified as playing a large and essential role in both hybrid male sterility and hybrid segregation distortion. We tested if the same genes cause hybrid sterility and hybrid segregation distortion by genetically dissecting the *sepia* region.

We generated 175 independent introgression lines in which the USA *sepia* region was moved into an otherwise pure Bogota background by backcrossing to the Bogota subspecies for 28 generations (14 of which involved recombination) (Fig. 1A) (22). The resulting introgression lines have genomes that derive almost entirely from Bogota, except for a small chromosomal region near *sepia*. We tested each introgression line for hybrid sterility and segregation distortion by crossing females from these introgression lines with USA males. The resulting hybrid males should be genetically identical to F₁ hybrid males, except for the small chromosomal region introgressed from USA. All 175 lines yielded hybrid males that were both fertile and produced normal (~50:50) sex ratios (Fig. 1B and figs. S1 and S2). This suggests that hybrid sterility and hybrid segregation distortion are caused either by

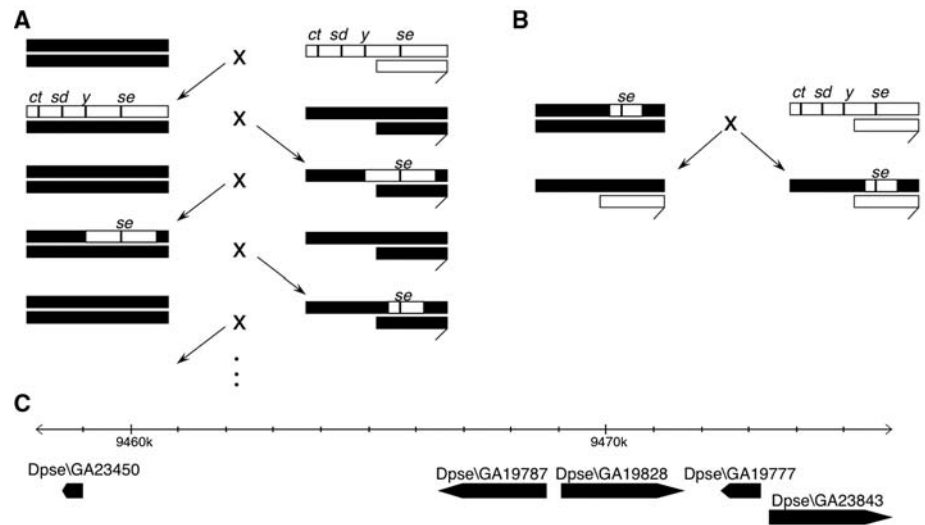


Fig. 1. (A) The *sepia* region of the USA X-chromosome was introduced into an otherwise pure Bogota background by using a crisscross design; recombination occurs in females, and the visible marker *sepia* was selected in males. Bogota material is black and USA material is white. Females have two X chromosomes, and males have one X chromosome and one Y chromosome (denoted by a hook). (B) When crossed with USA males, heterozygous introgression females produce two types of hybrid males. All *sepia* introgression hybrid males were fertile and showed normal segregation, whereas control F₁ hybrid males were mostly sterile and showed segregation distortion, as expected. (C) Hybrid sterility and segregation distortion both mapped to a region spanning ~20 kb that included five predicted genes, of which *GA19777* is the fastest evolving.

Department of Biology, University of Rochester, Rochester, NY 14627-0211, USA.

*To whom correspondence should be addressed. E-mail: pdns@mail.rochester.edu

tightly linked genes or by the same genes (or genes) in the region.

We genotyped all introgression lines using 41 microsatellite and 44 single-nucleotide polymorphism-based diagnostic markers [supporting online material (SOM) text]. The introgression breakpoints defined a region around *sepia* that harbors the hybrid sterility and hybrid segregation distortion genes (or genes). To isolate recombinants between the *sepia* locus and the hybrid sterility locus (recombinants that we did not recover above), we crossed Bogota females heterozygous for an introgression with USA males and tested *se*⁺ males for their ability to produce progeny. We found one such fertile *se*⁺ male. Analysis of flies derived from this recombinant showed that such males produce ~50:50 male:female progeny (fig. S3). Genotyping revealed that the gene (or genes) causing hybrid sterility and hybrid segregation distortion is located to the left of *sepia*. Together, the introgression and fertile *se*⁺ recombinant lines (which provided left and right breakpoints, respectively) localized the gene (or genes) causing hybrid sterility and hybrid segregation distortion to a ~20-kb region that contains five predicted genes (Fig. 1C).

No functional data are available from *Drosophila melanogaster* for the homologs of these genes. We sequenced the entire ~20-kb region from

the Bogota subspecies and compared it to the homologous USA sequence from the *D. pseudoobscura pseudoobscura* genome (23). We found no duplication, deletion, or obvious rearrangement differences between the subspecies in this region. But one of the predicted genes in the region, *GA19777*, shows eight nonsynonymous changes, which is surprising given its small coding region [591 base pairs (bp), excluding one 74-bp intron] (table S1). Given that genes causing reproductive isolation often evolve rapidly, *GA19777*, which contains a DNA-binding motif, represented our best candidate for the cause of hybrid sterility and/or hybrid segregation distortion.

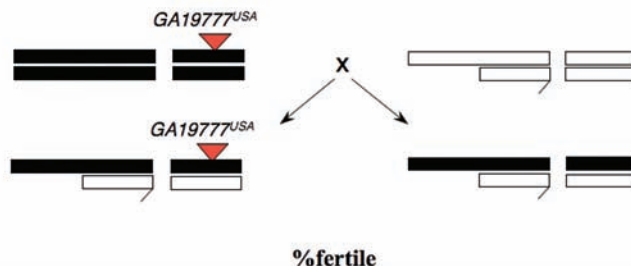
To determine whether *GA19777* causes these hybrid phenotypes, we first attempted to rescue the fertility of F₁ hybrid males with a transgenic copy of the (fertile) USA allele of *GA19777* (*GA19777*^{USA}). We cloned *GA19777*^{USA} along with ~0.2 kb of upstream and ~1.2 kb of downstream sequence into a *CaSpeR-4* P-element vector. The plasmid *P*{*w*⁺; *GA19777*^{USA}}, which carries a *D. melanogaster* miniwhite gene, was injected in Bogota-ER *white* embryos, and transformants were identified by wild-type eye color. When crossed with USA *white* males, Bogota females heterozygous for *P*{*w*⁺; *GA19777*^{USA}} produced two types of hybrid sons: those that inherited a transgenic *GA19777*^{USA} along with the endogenous Bogota

allele of *GA19777* (*GA19777*^{BOG}) (genotype *GA19777*^{BOG}/*Y*; *GA19777*^{USA}) and those that inherited only the endogenous *GA19777*^{BOG} allele (genotype *GA19777*^{BOG}/*Y*). The strength of fertility rescue in these experiments, if any, would thus depend on dominance relations between *GA19777*^{BOG} and the *GA19777*^{USA} transgene. Sperm motility assays showed that hybrid males that inherited the *GA19777*^{USA} transgene were more fertile than those that inherited only the endogenous *GA19777*^{BOG} regardless of the transgenic line studied (Fig. 2). We also crossed males with Bogota *white* females and found that hybrid males that inherited the *GA19777*^{USA} transgene produced progeny significantly more often than males that inherited only the endogenous *GA19777*^{BOG}, although fertility rescue was weak (Table 1). On the basis of these results, we conclude that *GA19777* plays a role in hybrid male sterility.

Surprisingly, *GA19777*^{USA} transgenic hybrids still produced almost all daughters (Table 1). *GA19777*^{USA} transgenes do not, therefore, suppress segregation distortion. This result can be explained in at least two ways: *GA19777* might cause hybrid sterility but not hybrid segregation distortion, or *GA19777* might cause both hybrid sterility and hybrid segregation distortion but each phenotype has different penetrance or degree of dominance (for example, with *GA19777*^{BOG} having more dominant effects on hybrid segregation distortion and more recessive effects on hybrid sterility).

To distinguish these possibilities, we performed a second transgenic experiment. In particular, we tested whether transgenic *GA19777*^{BOG} can induce segregation distortion. To do so, we cloned *GA19777*^{BOG} to the same length and coordinates as *GA19777*^{USA} and transformed USA and Bogota *white* strains with *P*{*w*⁺; *GA19777*^{BOG}}. As expected, no sterility or segregation distortion appeared in our transformed pure subspecies strains because such individuals lack the partner genes required for the hybrid incompatibility.

We tested whether *P*{*w*⁺; *GA19777*^{BOG}} can cause segregation distortion in hybrid sons produced by crossing introgression females to USA males. These sons carry all partner loci necessary for the appearance of hybrid problems (Bogota XL and USA autosomes) except for the Bogota factor (or factors) linked to *sepia* and are therefore fertile and nondistorting. We used females from a single introgression line (line104), which carries the smallest USA introgression near *sepia*. When crossed with USA males carrying *P*{*w*⁺; *GA19777*^{BOG}}, introgression females produced two types of hybrid sons: those that inherited a transgenic *GA19777*^{BOG} along with the endogenous *GA19777*^{USA} (genotype *GA19777*^{USA}/*Y*; *GA19777*^{BOG}) and those that inherited only endogenous *GA19777*^{USA} (genotype *GA19777*^{USA}/*Y*) (Fig. 3). Hybrid males that inherited the *GA19777*^{BOG} transgene produced a strikingly female-biased sex ratio when crossed with Bogota *white* females, whereas control males that inherited only endogenous *GA19777*^{USA} produced normal (~50:50) sex ratios (Fig. 3).



	%fertile		
USA1-1	74 (N =155)	25 (N =130)	P = 7.18E-55
USA1-2	91 (N =113)	24 (N =111)	P = 1.16E-62
USA1-7	79 (N =128)	20 (N =123)	P = 4.43E-67

Fig. 2. *P*{*w*⁺; *GA19777*^{USA}} rescues fertility in F₁ hybrid males, as measured blind in a sperm motility assay. Three independent transgenic strains were tested; the total number of males tested per genotype is denoted by *N*. The two genotypes are identical except for the presence or absence of the construct. *P* values were calculated using χ^2 tests with one degree of freedom. Red triangles indicate the inserted transgene.

Table 1. *P*{*w*⁺; *GA19777*^{USA}} rescues hybrid fertility but does not suppress segregation distortion. The transgene strains we used were the same as in Fig. 2. The two genotypes shown for each line are identical except for the presence or absence of the construct. Hybrid males producing one or more progeny (or no progeny) were classified as fertile (or sterile). We used a χ^2 test with one degree of freedom. SE values are shown in parentheses.

Line	Genotype	Fertile	Sterile	<i>P</i> value	Mean progeny	Females (%)
USA1-1	<i>w</i> ⁺	80	140		2.123 (0.33)	95.37 (1.5) <i>n</i> = 80
	<i>w</i>	7	174	2.37E-42	0.304 (0.13)	90.80 (6.91) <i>n</i> = 7
USA1-2	<i>w</i> ⁺	87	147		2.490 (0.42)	94.90 (1.10) <i>n</i> = 87
	<i>w</i>	0	254	1.60E-62	0.000	<i>n</i> = 0
USA1-7	<i>w</i> ⁺	56	146		1.327 (0.24)	94.77 (2.01) <i>n</i> = 56
	<i>w</i>	2	206	6.02E-58	0.024 (0.04)	75.00 (25.07) <i>n</i> = 2

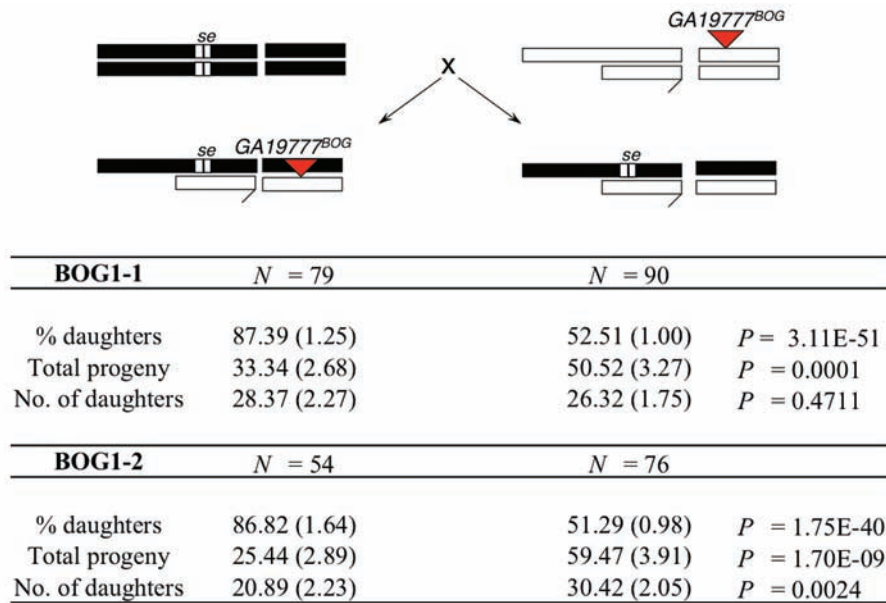


Fig. 3. $P\{w^+; GA19777^{BOG}\}$ induces segregation distortion in *sepia* introgression hybrid males. *Sepia* introgression hybrid males that inherit the $P\{w^+; GA19777^{BOG}\}$ transgene show significantly female-biased progeny as compared with those of males that do not inherit the transgene (when crossed with Bogota *white* females). The two genotypes are identical except for the presence or absence of the construct. Segregation distortion is also accompanied by a reduction in the total number of progeny and, in the case of BOG1-2, in the number of daughters produced. Two independent transgenic strains (BOG1-1 and BOG1-2) were tested; *N* denotes the number of fathers tested per genotype; mean values along with SE (in parentheses) are reported. *P* values were calculated using analysis of variance. Transgene males were inferred from the presence of w^+ individuals among F_1 and F_2 progeny. We modified the cross for the transgenic Bogota strain (BOG1-2): We first crossed introgression females with Bogota males carrying $P\{w^+; GA19777^{BOG}\}$ and collected female progeny. When crossed with USA *white* males, these females produced two types of hybrid sons: those that inherit a transgenic $GA19777^{BOG}$ along with endogenous $GA19777^{USA}$ and those that inherit only endogenous $GA19777^{USA}$. With both BOG1-1 and BOG1-2, transgenic hybrid males produced (rare) *white* and wild-type sons in roughly equal numbers, showing that sex ratio distortion is not due to transgene-induced male lethality.

Thus, *GA19777* plays a role in hybrid segregation distortion.

GA19777 is predicted to encode a polypeptide with a single Myb/SANT-like domain in Adf-1 (MADF) DNA-binding domain near its C terminus end. Reverse transcription polymerase chain reaction revealed that *GA19777* is expressed in the testes of both pure subspecies males and in sterile F_1 hybrid males (fig. S4). The constructs from the transgene experiments caused sex-chromosome segregation distortion and rescue of fertility even when located on autosomes, showing that transgenes have an effect in *trans* with respect to the sex chromosomes (table S2). *GA19777* therefore appears to act through its encoded product and not through *cis* chromosome effects. Further controls (fig. S5) showed that the segregation distortion/sterility effects of transgenes are not an artifact of gene dosage but depend on the species identity of the transgene.

The segregation distortion effect of $GA19777^{BOG}$ appears to reflect single-nucleotide substitutions that have occurred since the Bogota and USA subspecies split. In contrast, other known distorter genes involve gene duplications (24–26). The segregation distortion caused by *GA19777* is suppressed in pure Bogota individuals not be-

cause such individuals carry an insensitive Y chromosome [the Bogota Y can in fact be driven (21)] but because they are homozygous for recessive autosomal suppressors (21). Given its role in both segregation distortion (meiotic drive) and hybrid sterility, we named *GA19777 Overdrive (Ovd)*.

To study *Ovd*'s evolutionary history, we sequenced nine additional lines of Bogota and 16 of USA and found that seven of the eight nonsynonymous differences initially identified are fixed in the sample, whereas the remaining one (at position 61) is polymorphic within Bogota. Because Bogota strains lacking this segregating difference produce hybrids that are mostly sterile with segregation distortion (21), the change at position 61 does not seem essential for hybrid sterility or segregation distortion. We tested if *GA19777* experienced recurrent positive selection by performing a McDonald-Kreitman test (27); we also tested if *GA19777* experienced a recent selective sweep by calculating Tajima's *D* (28). These tests showed no significant departures from the null hypothesis of neutrality (SOM text). This result may reflect a lack of power due to unusually low polymorphism in both subspecies (SOM text), little divergence between these taxa, or possibly neutral evolution at *Ovd*.

Outgroup analysis with *D. miranda* and *D. persimilis* showed that all seven nonsynonymous changes and five synonymous changes fixed between Bogota and USA occurred in the Bogota lineage. Further analysis showed that rates of nonsynonymous and synonymous evolution are accelerated in the Bogota lineage relative to those expected given the species tree ($P = 7.7 \times 10^{-6}$ for nonsynonymous, $P = 0.0007$ for synonymous) (SOM text). A test for accelerated evolution [ratio of nonsynonymous/synonymous DNA changes (K_a/K_s)] in the Bogota lineage using PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>) was, however, nonsignificant ($P = 0.89$) (table S3), probably reflecting accelerated evolution at both nonsynonymous and synonymous sites in Bogota.

Our results show that the same gene, *Ovd*, affects both F_1 segregation distortion and hybrid sterility in evolutionarily young taxa that obey Haldane's rule. Although our results do not prove a history of segregation distortion in *D. pseudoobscura*, our results are consistent with and suggestive of such a history, especially given the burst of substitutions at *Ovd* in the Bogota lineage. Although evolutionary biologists usually explain the evolution of reproductive isolation as a side effect of adaptation to the ecological environment, our findings support the idea that genetic conflict, a form of adaptation to the internal genomic environment, may also be an important force in the evolution of postzygotic isolation.

References and Notes

- D. L. Hartl, *Theor. Popul. Biol.* **7**, 168 (1975).
- D. W. Hall, *Evolution* **58**, 925 (2004).
- T. W. Lyttle, *Annu. Rev. Genet.* **25**, 511 (1991).
- S. A. Frank, *Evolution* **45**, 262 (1991).
- L. D. Hurst, A. Pomiankowski, *Genetics* **128**, 841 (1991).
- B. Charlesworth, J. A. Coyne, H. A. Orr, *Genetics* **133**, 421 (1993).
- J. A. Coyne, H. A. Orr, *Evolution* **47**, 685 (1993).
- Y. Tao, D. L. Hartl, C. C. Laurie, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13183 (2001).
- S. Henikoff, K. Ahmad, H. S. Malik, *Science* **293**, 1098 (2001).
- D. A. Barbash, D. F. Siino, A. M. Tarone, J. Roote, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5302 (2003).
- J. Wittbrodt, R. Lammers, B. Malitschek, A. Ullrich, M. Scharl, *EMBO J.* **11**, 4239 (1992).
- C. T. Ting, S. C. Tsaur, M. L. Wu, C. I. Wu, *Science* **282**, 1501 (1998).
- D. C. Presgraves, L. Balagopalan, S. M. Abmayr, H. A. Orr, *Nature* **423**, 715 (2003).
- J. P. Masly, C. D. Jones, M. A. F. Noor, J. Locke, H. A. Orr, *Science* **313**, 1448 (2006).
- N. J. Brideau *et al.*, *Science* **314**, 1292 (2006).
- J. B. S. Haldane, *J. Genet.* **12**, 101 (1922).
- J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, Sunderland, MA, 2004).
- S. W. Schaeffer, E. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6097 (1991).
- S. Prakash, *Genetics* **72**, 143 (1972).
- H. A. Orr, S. Irving, *Genetics* **158**, 1089 (2001).
- H. A. Orr, S. Irving, *Genetics* **169**, 671 (2005).
- Materials and methods are available as supporting material on Science Online.
- S. Richards *et al.*, *Genome Res.* **15**, 1 (2005).
- C. Merrill, L. Bayraktaroglu, A. Kusano, B. Ganetzky, *Science* **283**, 1742 (1999).
- Y. Tao *et al.*, *PLoS Biol.* **5**, 2576 (2007).
- C. Montchamp-Moreau, D. Ogereau, N. Chaminade, A. Colard, S. Aulard, *Genetics* **174**, 1365 (2006).

27. J. H. McDonald, M. Kreitman, *Nature* **351**, 652 (1991).
 28. F. Tajima, *Genetics* **123**, 585 (1989).
 29. We thank C. Aquadro, J. Fry, R. Glor, H. Malik, D. Presgraves, Y. Tao, and the Orr lab for discussions; D. Lambert for advice on molecular techniques; H. Malik and N. Elde for help with PAML; and S. Schaeffer for fly stocks. Financial support was provided by NIH (GM51932). All sequences are available at GenBank

(accession numbers FJ349335 to FJ349342 and FJ418600 to FJ418631). The authors declare no conflict of interest.

Figs. S1 to S7
 Tables S1 to S9
 References

Supporting Online Material
www.sciencemag.org/cgi/content/full/1163934/DC1
 Materials and Methods
 SOM Text

29 July 2008; accepted 7 November 2008
 Published online 11 December 2008;
[10.1126/science.1163934](https://doi.org/10.1126/science.1163934)
 Include this information when citing this paper.

The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria

Nancy A. Moran,^{1*} Heather J. McLaughlin,¹ Rotem Sorek²

Among cellular organisms, symbiotic bacteria provide the extreme examples of genome degradation and reduction. However, only isolated snapshots of eroding symbiont genomes have previously been available. We documented the dynamics of symbiont genome evolution by sequencing seven strains of *Buchnera aphidicola* from pea aphid hosts. We estimated a spontaneous mutation rate of at least 4×10^{-9} substitutions per site per replication, which is more than 10 times as high as the rates previously estimated for any bacteria. We observed a high rate of small insertions and deletions associated with abundant DNA homopolymers, and occasional larger deletions. Although purifying selection eliminates many mutations, some persist, resulting in ongoing loss of genes and DNA from this already tiny genome. Our results provide a general model for the stepwise process leading to genome reduction.

Obligate symbionts and pathogens, which have evolved repeatedly from free-living bacterial ancestors, show striking convergence in fundamental genomic features. In several symbionts of insects, most ancestral genes are eliminated by deletion, resulting in some of the smallest known cellular genomes (1–4). Symbionts also display rapid evolution at both the DNA and peptide sequence levels and have highly biased nucleotide base com-

positions, with elevated frequencies of adenine and thymine (A+T). Because these genomes are asexual and do not acquire foreign DNA, each gene loss is irreversible (2, 5–7). These genomic features have been ascribed to increases in genetic drift associated with a host-restricted life-style (8, 9) and, potentially, to an increased mutation rate and biased mutational profile stemming from the loss of DNA-repair genes, which are among the gene categories most depleted in symbiont genomes (1, 10).

Although numerous sequenced examples of reduced genomes in obligate symbionts or pathogens are available, these are too distantly related to permit stepwise reconstruction of genomic changes. As a result, the dynamics of

ongoing genomic erosion, the extent to which mutation rate is elevated, the effectiveness of natural selection in purging mutations, and the nature of the mutational events that lead to further loss of DNA and metabolic functions are unclear. To illuminate these evolutionary processes, we sequenced several closely related genomes of the obligate symbiont *Buchnera aphidicola* from a single host species, the pea aphid *Acyrtosiphon pisum* (*Buchnera-Ap*). A previously sequenced genome of *Buchnera-Ap* showed a gene set typical for an obligate symbiont (1) lacking most ancestral genes, including genes underlying transcriptional regulation, biosynthesis of cofactors present in hosts, DNA repair, and other processes. The 607 retained genes encode machinery for replication, transcription, translation, and other essential processes, as well as biosynthetic pathways for essential amino acids required by hosts (1).

A. pisum is native to Eurasia, but has been introduced worldwide. It was first detected in North America in the 1870s (11). We sequenced the genomes of seven *Buchnera-Ap* strains descended from two colonizers of North America (and hence diverging up to 135 years ago), including two strains diverging in the laboratory for 7.5 years. Solexa sequencing was combined with verification by Sanger sequencing (12), to determine genomic sequences of these seven strains (Table 1). A total of 2392 positions (0.3% of sites on the 641-kb chromosome) showed a nucleotide substitution. These single-nucleotide polymorphisms (SNPs) were distributed approximately evenly around the chromosome (fig. S1). We also detected a total of 149 insertion or deletion events (indels): 134

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. ²Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

*To whom correspondence should be addressed. E-mail: nmoran@email.arizona.edu

Table 1. Description of sequence data.

	Tuc7	9-2-1	8-10-1	8-10-1	A2A	5AR	5A	7A
Source locality, year	Tucson AZ, 1999	Cayuga Co NY, 2001	Cayuga Co NY, 2001	Cayuga Co NY, 2001	Logan UT, 2003	Derived from 5A, 2000	Madison WI, 1999	Cayuga Co NY, 2000
Average read size	39	36	39	36	39	36	39	39
No. of initial reads	3,185,491	11,293,714	9,064,851	20,653,949	9,224,134	9,615,693	8,227,047	18,234,517
No. of reads mapped	1,024,330	6,731,726	4,432,760	12,977,253	7,088,978	6,944,135	4,448,342	12,150,323
Percent reads mapped*	32.16	59.61	48.90	62.83	76.85	72.22	54.07	66.63
Fold genome coverage (average)	57.9	369.9	259.4	700.5	412.2	374.6	256.0	661.1

*Unmapped reads represent contaminating DNA, largely from the host genome.