Title: Horizontal transmission and recombination maintain forever young bacterial symbiont genomes

Authors: Shelbi L. Russell[1,2*], Evan Pepper[2,3], Jesper Svedberg[2,3], Ashley Byrne[1], Jennie Ruelas Castillo[1], Christopher Vollmers[2,3], Roxanne A. Beinart[4], and Russ Corbett-Detig[2,3*]

Affiliations:

[1.] Department of Molecular Cellular and Developmental Biology. University of California Santa Cruz. Santa Cruz, California, United States of America.

[2.] Department of Biomolecular Engineering. University of California Santa Cruz. Santa Cruz, California, United States of America.

[3.] Genomics Institute, University of California, Santa Cruz, California, United States of America.

[4.] Graduate School of Oceanography. University of Rhode Island. Narragansett, Rhode Island, United States of America.

*To whom correspondence should be addressed: Email: shelbilrussell@gmail.com, russcd@gmail.com

Abstract:
When bacterial symbionts become associated with their hosts, their genomes are thought to decay inexorably towards an organelle-like fate due to decreased recombination and inefficient selection. Despite extensive theoretical treatment, no empirical study thus far has connected these underlying population genetic processes with long-term evolutionary outcomes. By sampling marine endosymbionts that range from primarily vertical to strictly horizontal transmission, we tested this canonical theory. We found that transmission mode strongly predicts recombination rates, and that exceedingly low recombination rates are associated with genome degradation in the marine symbionts with nearly strict vertical transmission. Nonetheless, even the most degraded marine endosymbiont genomes are occasionally horizontally transmitted and are much larger than their terrestrial insect symbiont counterparts. Therefore, horizontal transmission and recombination enable intermediate symbiont genome sizes that maintain substantial functional genetic variation.

Short title: Forever young bacterial symbiont genomes

One-sentence summary: Bacterial symbiont genomes maintain functional diversity through high rates of horizontal transmission and recombination.

Main Text:

Symbiont genome evolution theory predicts that upon host restriction, bacterial genomes begin the steady and inexorable process of decay due to decreased recombination rates, resulting in inefficacious natural selection (*1*). The transmission bottleneck that occurs when symbionts are passed on to host offspring further exacerbates these dynamics (*1*, *2*). In the early stages, deleterious mutations arise on each symbiont chromosome and a portion drift or hitch-hike with adaptive mutations to fixation. Subsequently, pseudogenes are lost entirely via deletion (*1*). Ultimately, this process is thought to result in an organelle-like genome that is a fraction of the size of its free-living ancestors. Although symbiont genome degradation is well-understood in theory (*e.g.*, (*1*–*3*)), no empirical study has directly evaluated the role of recombination in facilitating efficient natural selection and, thereby, the suppression of degradation. Interspecific-population level comparisons are essential for testing this long-standing question (*3*).

While the degraded genomes of many endosymbionts are consistent with this theory, such as symbionts of terrestrial sap and xylem-feeding insects (*4*), a diversity of associations present discrepancies. In particular, all known marine endosymbionts' genomes are at least one megabase in size and contain diverse coding content (*5*–*11*). This result is usually interpreted as reflecting the earliest stages of genome degradation (*12*, *13*), which is surprising considering the antiquity of these associations (*e.g.*, (*14*, *15*)). While important genes have been lost in some lineages, such as the recombination and repair gene *recA* and transversion mismatch-repair gene *mutY* (*16*), their loss did not portend complete genome degradation. Symbioses in marine environments exhibit significantly more horizontal transmission between hosts than those in terrestrial environments (*17*), suggesting that symbiont gene flow between hosts may prevent genome decay by enabling high rates of recombination, efficient natural selection, and the maintenance of highly diverse genome contents. This hypothesis has so far not been tested and it is not known whether marine endosymbionts represent the early stages of genome degradation as described by the canonical endosymbiont genome theory or if gene flow and recombination can stall such genome decay over evolutionary time.

To determine what evolutionary forces prevent marine endosymbiont genomes from degrading despite host restriction, we leveraged both population and comparative genomics of six marine bacterial-animal symbioses that exhibit modes of transmission across the spectrum from strict horizontal transmission (*Bathymodiolus* mytilid mussels), to mixed mode transmission (solemyid bivalves), to nearly strict vertical transmission (vesicomyid clams) (*17*) (see Fig. 1A,B, table S1). Each of these three groups evolved independently (fig. S1) to host intracellular gammaproteobacterial symbionts in their gills, which provide chemosynthetic carbon fixation to nutritionally support the association (*17*).

We used a combination of Nanopore and Illumina sequencing to assemble high quality reference genomes for the four previously uncharacterized endosymbionts and mitochondria for each host species ((*18*), table S2, species named in table S3). We paired these reference genomes with Illumina short-read data obtained from between seven and twenty

representatives of each population (table S1) to measure within population genetic variability ((*18*), tables S4, S5).

Comparative analyses of host mitochondrial and endosymbiont genealogies show strong evidence for horizontal transmission in all six populations (Fig 1C). Despite this apparent similarity in transmission mode, the genomes of vesicomyid endosymbionts are approximately one half the size of genomes of solemyid or mytilid endosymbionts, which themselves are approximately consistent with their free-living ancestors. Nonetheless, at 1-1.2 Mb, the partially degraded vesicomyid endosymbiont genomes are still ten times larger than the smallest insect endosymbionts (*4*).

Recombination is an important driver of genetic diversity in many bacterial populations (*19*, *20*). Although recombination can occur independently, horizontal transmission is required to create novel genetic combinations because recombination among clonal strains, *e.g.*, within a host-restricted clonal population of vertically transmitted endosymbionts, has little impact on haplotypic diversity (*21*). To explore the opportunity for recombination among genetically diverse symbionts, we partitioned symbiont genetic variation to between and within-host variation (*e.g.*, Fig 2A, tables S4, S5). Within vesicomyid symbionts, genetic diversity is strongly subdivided by hosts, and nearly all variation distinguishes host populations (Fig 2B). Conversely, for mytilid and solemyid endosymbionts, hosts have little impact, and two endosymbionts within a host are almost as divergent as two from different hosts (Fig 2B). These patterns are even more striking when examining the distribution of genetic diversity within host tissues. *Bathymodiolus* endosymbionts are exceptionally genetically diverse and the allele frequency spectra are similar to expectations for an equilibrium neutrally-evolving population (Fig 2C left, fig S2, S3), consistent with the high genetic diversities reported for other *bathymodiolin* symbionts (*22*). Conversely, solemyid endosymbionts maintain intermediate and more variable within-host genetic diversity, consistent with a mixture of vertical and horizontal transmission (Fig 2C middle, fig S4, S5). Finally, vesicomyid endosymbiont populations within hosts are virtually devoid of genetic variability (Fig 2C right, fig S6, S7). Therefore, despite their literal encapsulation within host tissues, mytilid and solemyid symbionts have abundant opportunity to recombine and create fitter chromosomes whereas vesicomyid symbionts must only rarely encounter genetically differentiated individuals.

Although opportunities are limited for vesicomyid endosymbionts, even relatively infrequent recombination events might drive patterns of genome evolution. We therefore developed a theoretical framework of symbiont evolution during mixed transmission modes. Importantly, our model demonstrates that these populations can be approximated using a standard Kingman-coalescent and that horizontal transmission is mechanisticanistically linked to observable recombination events ((*18*), figs S8, S9). We then performed extensive coalescent simulations and used a machine-learning regression framework to estimate the effective recombination rates for each population ((*18*), table S6). Estimated recombination rates are substantially higher in mytilid and solemyid symbionts than in vesicomyid symbionts (Fig 2D, table S4, table S7), indicating that the potential for recombination within hosts is realized in

these species. In fact, these populations have some of the largest effective recombination rates ever reported for bacteria (*23*, *24*). Nonetheless, we also found modest rates of recombination within both partially degenerate vesicomyid genomes, despite the loss of many requisite genes in the *Calyptogena magnifica* symbiont (*16*, *25*).

A fundamental consequence of decreased recombination rates for endosymbionts is that selected mutations cannot be shuffled to form higher fitness chromosomes. Ultimately, this competition among selected mutations on different haplotypes, clonal interference, can drive the fixation of deleterious mutations and reshape genealogies towards long terminal branches and excesses of rare alleles (*26*). Consistent with theory, we find abundant rare alleles in the vesicomyid symbiont genomes (Fig 3, fig S10, table S4; Tajima's D = -1.98 and -2.03 for *C. magnifica* and *Calyptogena fausta*, respectively), but little skew in the allele frequencies of other endosymbiont populations (D ranges from 0.196 to -1.01). Importantly, it is unlikely that differences in host species demography have driven these differences, *e.g.,* recent population expansions specifically in vesicomyid clams. In fact, we found less allele frequency skew in the mitochondrial genomes than in the vesicomyid symbionts for all species considered, and the strongest negative skew in the allele frequencies in the mitochondrial genome of *Bathymodiolus septemdierum* from the Lau Basin (D = -1.9, table S4). Additionally, weakly selected fixations between symbiont populations follow the expected trend, with dN/dS values of 0.14, 0.096, 0.083 for vesicomyid, solemyid and bathymodiolin genomes, respectively (pairwise p-values: bathymodiolin-vesicomyid p=4.60e-14, solemyid-vesicomyid p= 1.02e-11, and bathymodiolin-solemyid p= 0.0365), consistent with recombination enabling more efficacious selection for sustained periods of time.

Genome structure stasis is thought to be another hallmark of canonical endosymbiont genome evolution, and many terrestrial endosymbioses have reported static degenerate genomes (*4*). Whole genome alignments reveal that the vesicomyid symbiont genomes are highly syntenic, with few rearrangements, insertions, or deletions; whereas the other symbiont genomes are highly structurally dynamic (Fig 4). It is, therefore, likely that strong clonal interference has driven genome erosion in these groups. This could proceed by iteratively locking deletions (initially caused by recombination (*25*), as depicted in Fig 4A) together through many generations of symbiont transmission bottlenecks, incrementally reducing the size of the genome. With recombination events occurring so rarely in these endosymbiont genomes, natural selection is unable to efficiently purge deleterious deletions. Inversions, which can be highly mutagenic by inverting the translated strand, inducing replication-transcription machinery collisions (*27*), may be nearly absent due to their high fitness costs.

In contrast, the solemyid symbionts are as structurally dynamic as strictly horizontally transmitted associations, and they have far more mobile genetic elements than any of the other symbiont genomes (table S8). This is consistent with a combination of free-living and host associated periods (*28*), and mirrors the early stages of host association (*12*, *13*), as well as the early stages of eukaryotic asexuality (*29*). Many of these elements exhibit homology to different environmental bacteria, implying many independent insertion events (table S9).

It is plausible that vesicomyid symbionts' horizontal transmission and recombination rates are the beginning of the range of values that permit genome decay. They undergo horizontal transmission (Figs 1C, 2B) and recombine (Fig 2D, table S4). However, the signatures of clonal interference and sustained intermediate genome sizes indicate that selection is sufficiently efficacious to maintain some functional diversity in these populations, counter to the expectations for the original theory on endosymbiont genome evolution (*1*, *2*).

The extreme age of these symbioses and the conservation of evolutionary signatures across groups suggest that patterns in genome structure are maintained over long periods of time. For example, symbiotic vesicomyids evolved from their non-symbiotic vesicomyid ancestors around 66 mya (*14*) and the solemyids evolved from their non-symbiotic protobranch ancestors even earlier, with estimates dating back to more than 400 mya (*15*). While host-switching and novel symbiont acquisition have certainly occurred, phylogenetic evidence suggests it has been relatively rare (*9*, *30*). Thus, despite their old age, necessity for host survival, and transmission route through host oocytes, the symbiont genomes in these associations have not degraded to the minimal sizes seen in insect symbionts. Instead, moderate rates of recombination have allowed the solemyid and vesicomyid symbiont genomes to maintain functional diversity characteristic of a free-living or moderately reduced genome, respectively.

Here we show empirically that symbiont genome sizes and functional diversity are predicted by the rate of gene flow into and among symbiont populations via horizontal transmission and recombination. These results suggest that there is a range of possible intermediate genome degradation states that can be maintained indefinitely, even over millions of years, with sufficient recombination. Therefore, symbiont genome evolution following host restriction is not a one-way, inescapable process that ends in an organelle-like state as it is commonly presented (*1*, *2*). These results validate long-standing but untested theory and demand a reinterpretation of the vast diversity of symbioses that exhibit intermediate rates of horizontal transmission and incomplete genome degradation.

References and Notes:

1. J. P. McCutcheon, N. A. Moran, *Nat. Rev. Microbiol.* (2011), doi:10.1038/nrmicro2670.
2. C. Toft, S. G. E. Andersson, *Nat. Rev. Genet.* **11**, 465–475 (2010).
3. J. J. Wernegreen, *Ann. N. Y. Acad. Sci.* **1360**, 16–35 (2015).
4. N. A. Moran, G. M. Bennett, *Annu. Rev. Microbiol.* **68**, 195–215 (2014).
5. I. L. G. Newton *et al.*, *Science*. **315**, 998–1000 (2007).
6. H. Kuwahara *et al.*, *Curr. Biol.* **17**, 881–886 (2007).
7. O. Dmytrenko *et al.*, *BMC Genomics*. **15** (2014), doi:doi:10.1186/1471-2164-15-924.
8. I. J. Miller, N. Vanee, S. S. Fong, G. E. Lim-Fong, J. C. Kwan, *Appl. Environ. Microbiol.* **82**, 6573–6583 (2016).
9. S. L. Russell, R. B. Corbett-Detig, C. M. Cavanaugh, *ISME J.*, 1359–1371 (2017).
10. T. A. Hendry *et al.*, *mBio*. **9**, e01033-18, /mbio/9/3/mBio.01033-18.atom (2018).
11. O. Jäckle *et al.*, *Proc. Natl. Acad. Sci.*, 201818995 (2019).
12. L. Ran *et al.*, *PLoS ONE*. **5**, e11486 (2010).

13. K. F. Oakeson *et al.*, *Genome Biol. Evol.* **6**, 76–93 (2014).
14. S. B. Johnson, E. M. Krylova, A. Audzijonyte, H. Sahling, R. C. Vrijenhoek, *Syst. Biodivers.* **15**, 346–360 (2017).
15. P. P. Sharma *et al.*, *Mol. Phylogenet. Evol.* **69**, 188–204 (2013).
16. H. Kuwahara *et al.*, *BMC Evol. Biol.* **11**, 285 (2011).
17. S. L. Russell, *FEMS Microbiol. Lett.*, fnz013 (2019).
18. Materials and methods are available as supplementary materials at the Science website.
19. B. J. Shapiro *et al.*, *Science*. **336**, 48–51 (2012).
20. M. J. Rosen, M. Davison, D. Bhaya, D. S. Fisher, *Science*. **348**, 1019–1023 (2015).
21. R. A. Chong, H. Park, N. A. Moran, *Mol. Biol. Evol.* **36**, 1481–1489 (2019).
22. R. Ansorge *et al.*, *bioRxiv* (2019), doi:10.1101/531459.
23. M. Vos, X. Didelot, *ISME J.* **3**, 199–208 (2009).
24. M. Lin, E. Kussell, *Nat. Methods*. **16**, 199–204 (2019).
25. A. I. Nilsson *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12112–12116 (2005).
26. R. A. Neher, *Annu. Rev. Ecol. Evol. Syst.* **44**, 195–215 (2013).
27. C. N. Merrikh, H. Merrikh, *Nat. Commun.* **9**, 4662 (2018).
28. I. L. G. Newton, S. R. Bordenstein, *Curr. Microbiol.* **62**, 198–208 (2011).
29. S. Glémin, N. Galtier, in *Evolutionary Genomics*, M. Anisimova, Ed. (Humana Press, Totowa, NJ, 2012; http://link.springer.com/10.1007/978-1-61779-582-4_11), vol. 855, pp. 311–335.
30. G. Ozawa *et al.*, *Genome Biol. Evol.* **9**, 2226–2236 (2017).
31. R. Ponnudurai *et al.*, *Stand. Genomic Sci.* **12** (2017), doi:10.1186/s40793-017-0266-y.
32. J. Ruan, H. Li, “Fast and accurate long-read assembly with wtdbg2” (preprint, Bioinformatics, 2019), , doi:10.1101/530972.
33. B. J. Walker *et al.*, *PLoS ONE*. **9**, e112963 (2014).
34. H. Li, *arXiv*, in press.
35. Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin, *Bioinformatics*. **28**, 1420–1428 (2012).
36. A. Bankevich *et al.*, *J. Comput. Biol.* **19**, 455–477 (2012).
37. D. Hyatt, P. F. LoCascio, L. J. Hauser, E. C. Uberbacher, *Bioinformatics*. **28**, 2223–2230 (2012).
38. C. Camacho *et al.*, *BMC Bioinformatics*. **10**, 421 (2009).
39. K. Lagesen *et al.*, *Nucleic Acids Res.* **35**, 3100–3108 (2007).
40. T. M. Lowe, S. R. Eddy, *Nucleic Acids Res.* **25**, 955–964 (1997).
41. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Res.* **25**, 1043–1055 (2015).
42. M. Wu, J. A. Eisen, *Genome Biol.* **9**, 1 (2008).
43. D. R. Kelley, M. C. Schatz, S. L. Salzberg, *Genome Biol.* **11**, 1 (2010).
44. M. Bernt *et al.*, *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
45. H. Liu, S. Cai, H. Zhang, R. C. Vrijenhoek, *Mitochondrial DNA Part A*. **27**, 4333–4335 (2016).
46. F. Plazzi, A. Ribani, M. Passamonti, *BMC Genomics*. **14**, 1 (2013).
47. H. Li *et al.*, *Bioinformatics*. **25**, 2078–2079 (2009).
48. M. A. DePristo *et al.*, *Nat. Genet.* **43**, 491–498 (2011).
49. P. Danecek *et al.*, *Bioinformatics*. **27**, 2156–2158 (2011).
50. S. L. Russell, C. M. Cavanaugh, *Mol. Biol. Evol.* **34**, 2747–2761 (2017).
51. A. Stamatakis, *Bioinformatics*. **30**, 1312–1313 (2014).
52. S. Kurtz *et al.*, *Genome Biol.* **5**, 1 (2004).
53. T. Ikuta *et al.*, *ISME J.* **10**, 990 (2016).

54. R. Leplae, G. Lima-Mendez, A. Toussaint, *Nucleic Acids Res.* **38**, D57–D61 (2010).
55. D. Bi *et al.*, *Nucleic Acids Res.* **40**, D621–D626 (2012).
56. V. Ranwez, S. Harispe, F. Delsuc, E. J. P. Douzery, *PLoS ONE*. **6**, e22594 (2011).
57. Z. Yang, *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
58. A. E. Darling, B. Mau, N. T. Perna, *PLoS ONE*. **5**, e11147 (2010).
59. D. W. Waite *et al.*, *Front. Microbiol.* **9**, 772 (2018).
60. K. D. Yamada, K. Tomii, K. Katoh, *Bioinformatics*. **32**, 3246–3251 (2016).
61. M. Kearse *et al.*, *Bioinformatics*. **28**, 1647–1649 (2012).
62. J. Wakeley, N. Aliacar, *Genetics*. **159**, 893–905 (2001).
63. J. R. Pannell, B. Charlesworth, *Philos. Trans. R. Soc. B Biol. Sci.* **355**, 1851–1864 (2000).
64. J. Wakeley, *Mol. Ecol.* **13**, 865–875 (2004).
65. K. Komaki, H. Ishikawa, *Insect Biochem. Mol. Biol.* **30**, 253–258 (2000).
66. M. S. Wollenberg, E. G. Ruby, *Appl. Environ. Microbiol.* **75**, 193–202 (2009).
67. C. M. Cavanaugh, Z. P. McKiness, I. L. G. Newton, F. J. Stewart, in *The Prokaryotes*, M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. Stackebrandt, Eds. (Springer New York, New York, NY, 2006; http://link.springer.com/10.1007/0-387-30741-9_18), pp. 475–507.
68. J. Klose *et al.*, *Proc. Natl. Acad. Sci.* **112**, 11300–11305 (2015).
69. S. Duperron, A. Quiles, K. M. Szafranski, N. Léger, B. Shillito, *Front. Mar. Sci.* **3**, 16 (2016).
70. R. Sender, S. Fuchs, R. Milo, (2016) (available at http://biorxiv.org/lookup/doi/10.1101/036103).
71. T. Ikuta *et al.*, *R. Soc. Open Sci.* **3**, 160130 (2016).
72. S. L. Russell, E. McCartney, C. M. Cavanaugh, *Proc. R. Soc. B Biol. Sci.* **285**, 9 (2018).
73. M. Kaltenpoth, W. Goettler, S. Koehler, E. Strohm, *Evol. Ecol.* **24**, 463–477 (2010).
74. W. Z. Stephens *et al.*, *mBio*. **6**, e01163-15 (2015).
75. X. Didelot, A. S. Walker, T. E. Peto, D. W. Crook, D. J. Wilson, *Nat. Rev. Microbiol.* **14**, 150–162 (2016).
76. A. Mira, N. A. Moran, *Microb. Ecol.* **44**, 137–143 (2002).
77. M. Möhle, *Adv. Appl. Probab.* **30**, 493–512 (1998).
78. M. Möhle, M. Notohara, *J. Appl. Probab.* **53**, 953–956 (2016).
79. S. Matuszewski, M. E. Hildebrandt, G. Achaz, J. D. Jensen, *Genetics*. **208**, 323–338 (2018).
80. M. Möhle, S. Sagitov, *Ann. Probab.* **29**, 1547–1562 (2001).
81. B. Eldon, J. Wakeley, *Genetics*. **181**, 615–629 (2008).
82. J. P. Wares, C. W. Cunningham, (2001).
83. M. Lynch *et al.*, *Proc. Natl. Acad. Sci.* **105**, 9272–9277 (2008).
84. C. Haag-Liautard *et al.*, *PLoS Biol.* **6**, e204 (2008).
85. K. Yahara *et al.*, *Mol. Biol. Evol.* **33**, 456–471 (2016).
86. P. D. Dixit, T. Y. Pang, S. Maslov, *Genetics*. **207**, 281–295 (2017).
87. J. Kelleher, A. M. Etheridge, G. McVean, *PLOS Comput. Biol.*, 22 (2016).
88. L. Excoffier, M. Foll, *Bioinformatics*. **27**, 1332–1334 (2011).
89. N. De Maio, D. J. Wilson, *Genetics*. **206**, 333–343 (2017).
90. G. A. Watterson, *Theor. Popul. Biol.* **7**, 256–276 (1975).
91. M. Nei, W.-H. Li, *Proc. Natl. Acad. Sci.* **76**, 5269–5273 (1979).
92. K. Polimis, A. Rokem, B. Hazelton, *J. Open Source Softw.* **2**, 124 (2017).

Acknowledgements:
We thank Colleen Cavanaugh and Peter Girguis for *Calyptogena* freezer samples collected previously. For their assistance in collecting *Bathymodiolus* samples from the Lau Basin and Veatch Canyon, respectively, we thank the crews of the R/V *Falkor* and ROV *ROPOS* and the R/V *Atlantis* and HOV *Alvin*. **Funding:** this work was supported by UC Santa Cruz, Harvard University, the Alfred P. Sloan Foundation, the NIH (R35GM128932), the NSF (OCE-1819530), the Schmidt Ocean Institute, and the UNOLS Early Career Training Cruise Program. **Author Contributions:** SLR, study conception and design, samples, Illumina data production, analysis, and writing; EP, Illumina and Nanopore data production; JS, analysis of genome structures and writing; AB, Nanopore data production; JRC, assistance with sampling; CV, Nanopore data production; RB, samples and intellectual contributions; RCD, study conception and design, Illumina and Nanopore data production, theory, analysis, and writing. Competing interests: none. Data and materials availability: data generated in this study is available through NCBI BioProject number PRJNA562081 and code written to analyze it is available from https://github.com/shelbirussell/ForeverYoungGenomes_Russell-et-al.

List of Supplementary Materials:
Materials and Methods
figs S1-S10
tables S1-S10
References (32-92 only cited in the supplemental)

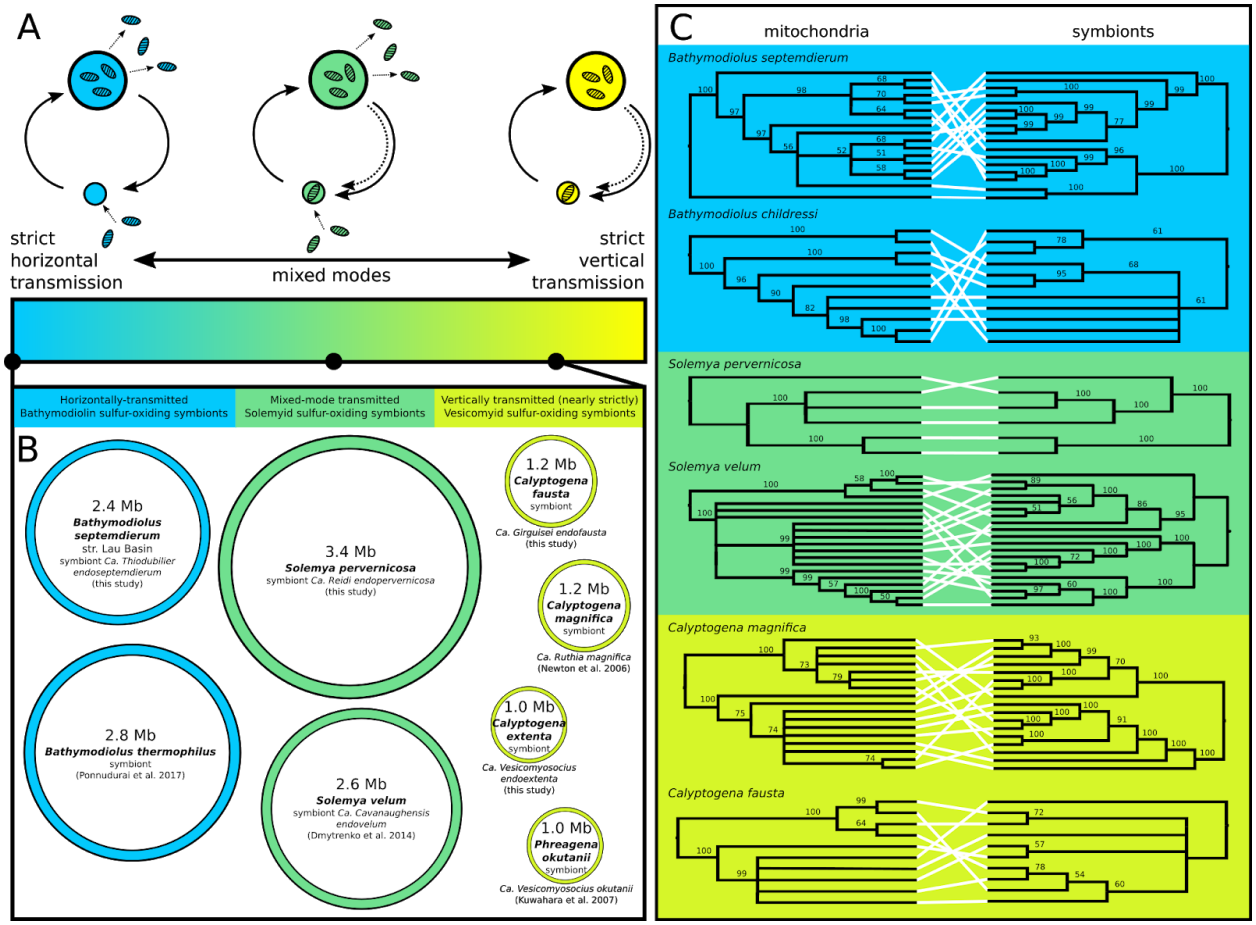

Fig. 1. Nearly-strictly vertically transmitted chemosynthetic endosymbionts exhibit genome erosion despite ongoing horizontal transmission events in their populations. A) Transmission mode spectrum from strict horizontal transmission to strict vertical transmission, with a diversity of mixed modes, incorporating both strategies, in between. B) Genome sizes from this and previous studies (*5–7*, *31*) reveal consistent patterns of genome erosion among the vesicomyid symbionts, but not the other groups with higher rates of horizontal transmission. C) Mitochondrial and symbiont genealogies are discordant for all groups, indicating that sufficient amounts of horizontal transmission occur in vesicomyid populations to erode the association between these cytoplasmic genomes. Maximum likelihood cladograms with nodes below 50% bootstrap support collapsed. Species are color coded by their symbiont transmission mode as in A).

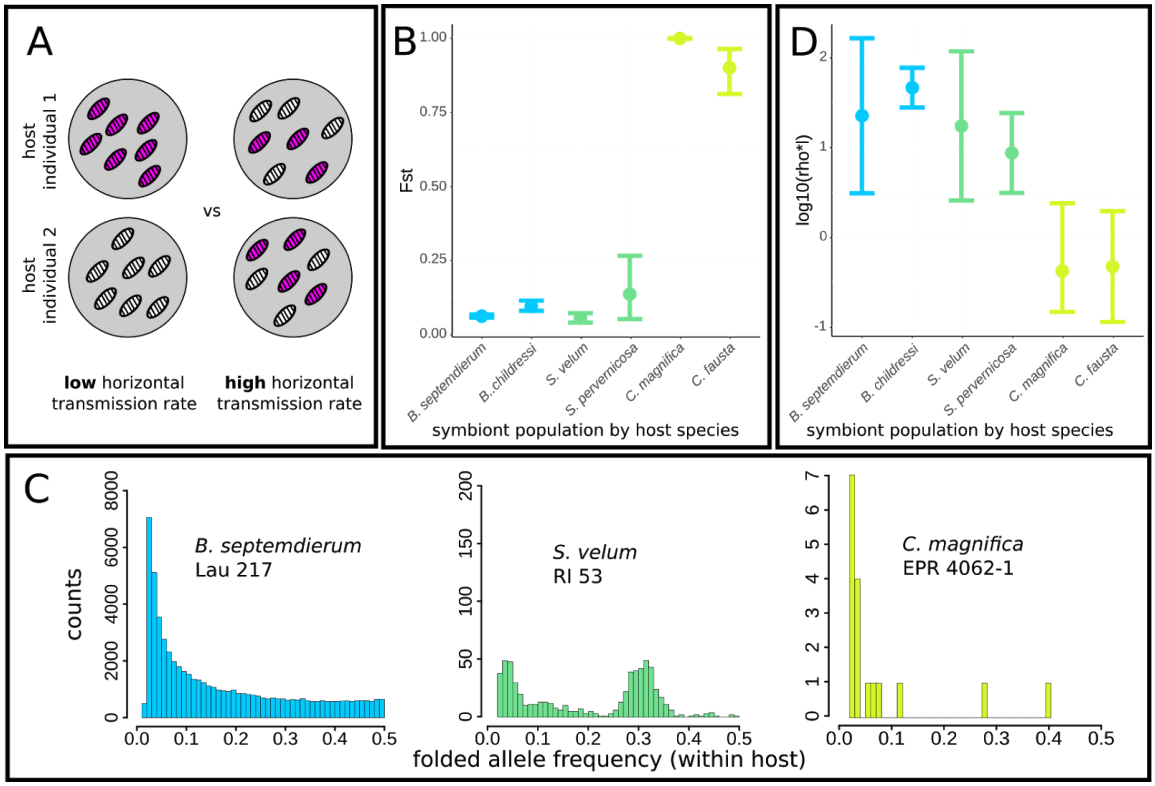

Fig. 2. Horizontal transmission and recombination introduce novel alleles into symbiont populations. A) Model of endosymbiont genotype (pink vs. white) distributions under well-mixed high horizontal transmission rates (top) and differentiated low horizontal transmission rates. B) Horizontally transmitted mytilid (blue) and mixed mode transmitted solemyid (green) symbionts are well mixed among hosts, whereas the nearly strictly vertically transmitted vesicomyid symbionts (yellow-green) are highly differentiated among hosts. Error bars = 95% confidence intervals from non-parametric bootstrapping. C) Intrahost population folded allele frequency spectra (AFS) are shaped by access to gene flow, which is enabled by horizontal transmission and recombination. D) Recombination rates are significantly higher in the mytilid (blue) and solemyid (green) symbiont genomes compared to the vesicomyid symbiont genomes (yellow-green). Error bars = 95% confidence intervals.

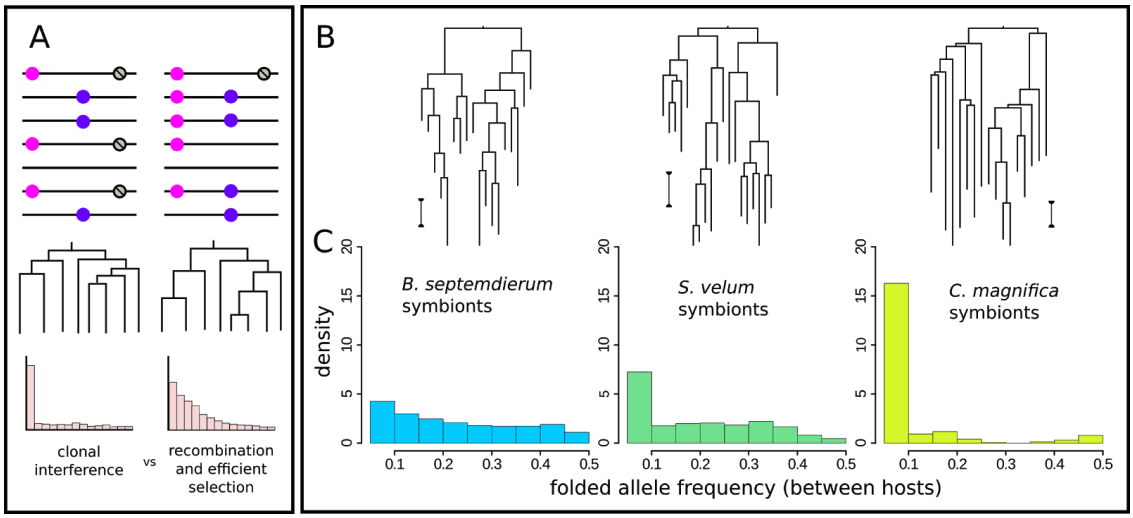

Fig. 3. Consequences of access to gene flow via horizontal transmission and recombination on the distribution of symbiont genetic diversity between hosts. A) Diagram showing how beneficial alleles (pink) are linked to deleterious alleles (grey) in populations experiencing clonal interference versus free recombination, and how these processes are reflected in the underlying population genealogies and allele frequency spectra. B) Symbiont genealogies as labeled in C). Scale bars from left to right: 6e-4, 3e-5, 8e-6 substitutions/site. C) Between host allele frequency spectra (AFS) for three species representing the three groups of hosts.

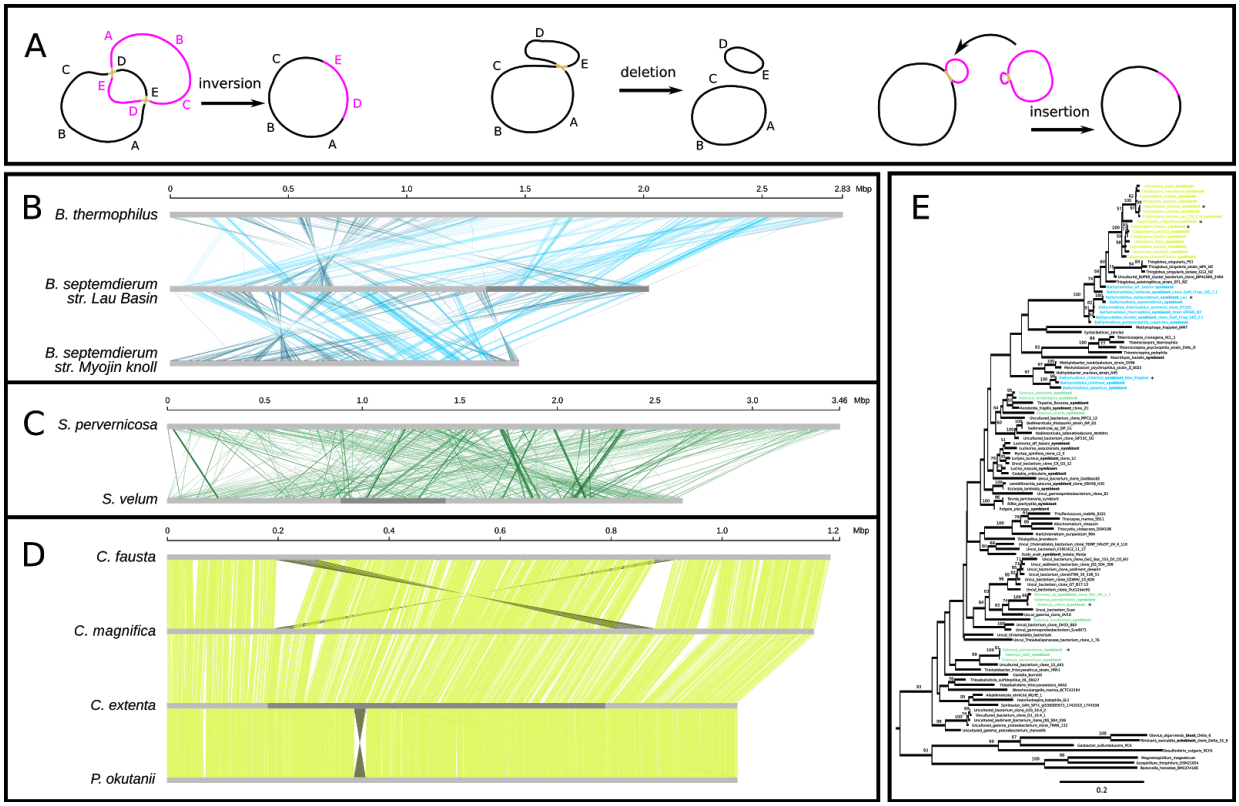

Fig. 4. Genome structure is shaped by horizontal transmission and recombination. A) Models of recombination-based structural mutation mechanisms. B-D) Whole genome alignments for B) mytilid, C) solemyid, and D) vesicomyid symbiont genome assemblies with >1 Mb scaffolds. E) Maximum likelihood 16S rRNA phylogeny for the chemosynthetic gammaproteobacterial symbionts and their free-living relatives (taxa in table S10).