Genomic Data Analyses Unveil Cases of Polygenic Evolution

Powerful analytic tools are beginning to establish how changes involving suites of genes lead to new traits in evolving microbes

Rachel B. Brem

Short-read sequencing is revolutionizing our understanding of how microbes evolve. Scientists working in academic laboratories now routinely sequence the genomes or transcriptomes of hundreds of microbial isolates, a prelude to addressing questions about what those data tell us. For evolutionary biologists, a central goal is to understand how new phenotypes arise in response to environmental changes. The answer to such questions lies somewhere among the molecular profiles of wild strains, including the genes that enable pathogens to adapt to hosts or halophiles to high-salt conditions.

The big challenge is that genomes are littered with changes that have little or no interesting evolutionary history. By genetic drift alone, replicating microbes pass to their progeny plenty of mutations that do not confer any advantages along with some mutations that likely reduce fitness. Because these nonadaptive variants accumulate along with adaptive variants in genomes, they vastly outnumber the loci that can teach us how new traits arise during evolution. This signal-to-noise issue drives research in evolutionary genetics in organisms up as well as across the tree of life.

Simple and Polygenic Adaptations

To look for cases of natural selection in molecular data, geneticists keep in mind two models of evolutionary innovation. According to one of these models, mutations in a single gene can give rise to a major reproductive advantage, leaving signatures in the genome that can be recognized computationally. This simple picture continues to motivate the great majority of molecular-evolution studies of wild microbes that have scientists searching painstakingly for beneficial changes in one or a few genes at a time.

According to the other model, combinations of multiple variants, in genes all over the genome, could together give rise to adaptive traits. Each locus could confer a partial benefit, making incremental contributions to single phenotypes. Alternatively, a complete set of changes might be required all at once for an organism to develop a beneficial behavior.

In the latter polygenic scenarios, the signature of natural selection at any one adaptive gene can be undetectably subtle on its own. However, with panels of genome sequence data or of transcriptomes of wild microbes, investigators can combine data across genes and genomes, bringing otherwise weak signals to the fore. Several new analytic tools are sufficiently powerful to begin to establish how changes involving suites of genes lead to new traits in evolving microbes.

SUMMARY

➤ Although combinations of multiple variants in genes all over the genome could give rise to adaptive traits, identifying evidence for those changes presents a major signal-to-noise challenge.
➤ By combining sequence data across genes and genomes, investigators can bring otherwise weak signals of polygenic evolution to the fore.
➤ Suites of multiple mutations, identifiable as groups across bacterial populations, can help particular strains of Burkholderia dolosa to colonize specific patients.
➤ Transfers of sets of related genes, each at a different point in evolutionary history, would not be expected by chance and suggests a drive toward improved fitness.
➤ Similarly, when many separate mutations affect expression of genes of similar function, they appear to attest to an importance for fitness.
Sequence Tests for Polygenic Evolution

The search for polygenic evolution starts with tests for natural selection that make use of DNA sequence data. Some of these methods look for patches of allele sharing across strains, which might be expected when a beneficial mutation sweeps through a population of sexually recombining individuals. Others look for multiple interspecies changes at a particular gene, suggestive of ancient, repeated evolutionary tinkering at that locus in one lineage relative to another.

For the most part, these tests were developed for analyzing one locus at a time. But pooling the results across genes and loci can enable investigators to detect evidence for polygenic evolution. Sequence signatures of natural selection can and do show up in multigene sets of annotated drug resistance or virulence factors, for example, or multiple genes encoding cell-surface proteins that interact with the environment.

Analysis of clinical isolates of *Burkholderia dolosa* bacteria from cystic fibrosis patients uncovered a set of 17 genes with recurrent mutations across strains, in which protein-coding changes are far more prevalent than are silent mutations (Fig. 1), according to Roy Kishony at Harvard University in Cambridge, Mass., and his collaborators. No single strain genome contained all the variants, and no single mutation was present in all strains. A compelling interpretation is that distinct combinations of these loci, which include known pathogenicity genes as well as some surprises, confer advantages to each strain as it adapts to conditions encountered in individual hosts.

In this case, fitness would not be attributed to any one mutation on its own. Instead, a suite of

---

**FIGURE 1**

<table>
<thead>
<tr>
<th>Function</th>
<th>m</th>
<th>Patient</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>O$_2$-related</td>
<td>17</td>
<td></td>
<td>histidine kinase <em>fixL</em></td>
</tr>
<tr>
<td>gene regulation</td>
<td></td>
<td></td>
<td>response regulator <em>fixJ</em></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>transcriptional regulator <em>frn</em></td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
<td></td>
<td>methyltransferase</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>sigma factor</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>glucoamylase</td>
</tr>
<tr>
<td>Antibiotic resistance</td>
<td>14</td>
<td></td>
<td>ribosomal protein L4 <em>rpl4</em></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td></td>
<td>DNA gyrase subunit <em>gyrA</em></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>beta-lactamase <em>bla1</em></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td>ABC-type transporter <em>mdlB</em></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>DNA gyrase subunit B <em>gyrB</em></td>
</tr>
<tr>
<td>Outer-membrane synthesis</td>
<td>10</td>
<td></td>
<td>glycosyltransferase <em>wbaD</em></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>glycosyltransferase <em>mfaA</em></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>transporter <em>ipIT</em></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>permease <em>iptG</em></td>
</tr>
<tr>
<td>Secretion</td>
<td>4</td>
<td>A B C D E F G H I J K L M N</td>
<td>exoprotein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>type II secretion <em>gspD</em></td>
</tr>
</tbody>
</table>

Each of 17 *B. dolosa* genes (rows) harbored an acquired mutation in several human subjects, as signified by squares (color intensity indicates the number of mutations observed among isolates from a given subject). The total number of mutations observed within each gene, m, is indicated at left. Genes are grouped by biological function and labeled with the annotations of close homologs and, when available, with the names of these homologs. (Adapted with permission from T. D. Lieberman et al., Nature Genet. 43:1275–1280, 2011.)
multiple mutations, identifiable as a group across the population, helped each *Burkholderia* isolate to colonize a human patient.

**Polygenic Evolution by Horizontal Gene Transfer**

Microbial genomes also lend themselves well to tests for polygenic evolution via horizontal gene transfer (HGT), the exchange of genetic material from one species to another. Because HGT events are random occurrences, a single transferred locus in a microbial genome cannot indicate anything about its potential benefit to the organism. However, transfer of a set of related genes, each at a different point in evolutionary history, would not be expected by chance and suggests a drive toward improved fitness.

For example, the genome of *Encephalitozoon hellem*, a microsporidian animal pathogen, stands out from other related species by virtue of its containing a suite of folate biosynthesis genes, according to Patrick Keeling of the University of British Columbia (BC) in Vancouver, BC, Canada, and his collaborators. Some of these genes were acquired from bacteria, some from fungi, and some from metazoans (Fig. 2). The evolutionary gene collecting appears specific to folate metabolism, in that other pathways in *E. hellem* show the usual sequence patterns of vertical inheritance from a microsporidian ancestor.

The likely evolutionary picture is one in which *E. hellem* assembled its folate genes by multiple HGT events over time, with the resulting pathway conferring a fitness advantage—perhaps during infections of folate-deficient hosts.

**Expression Tests for Polygenic Evolution**

Apart from DNA sequence-based tests, mRNA expression data can also point to cases of polygenic evolution in microbes. Here again, analysis of multiple independent genetic changes can provide a leg up on the study of any one locus alone. A single regulatory mutation in a species does not necessarily reflect a history of natural selection, even in the case of, for instance, a transcription factor affecting several gene targets. However, when many separate mutations affect expression of genes of similar function, they appear to attest to an importance for fitness.

In the yeast *Saccharomyces paradoxus*, regulatory information at each of dozens of membrane protein stress-response genes encodes a new, high-expression program in rich medium, according to Hilary Martin, now at the Wellcome Trust Centre for Human Genetics, and other collaborators of mine at the University of California, Berkeley. A descent into molecular disrepair does not explain those changes, particularly because promoter sequences of the membrane protein genes are tightly conserved among individual cells of *S. paradoxus*.

Natural selection is the likely basis for the species-specific expression pattern. Perhaps the up-regulating alleles boost membrane protein levels and, plausibly, protect *S. paradoxus* against environmental stress.

**The Rationale for Genetic Complexity**

Taken together, these results make clear that strategies are in hand for studying polygenic evolution through use of molecular data. When applied to microbial genomes and transcriptomes, these methods can be used to identify strong signals of natural selection. In silico results aside, though, biologists need a convincing rationale before accepting the idea of complex, polygenic evolutionary changes. How could a suite of independent mutations lead to the development of a new behavior in a microorganism if one mutation could suffice?

One reason could be that it is rare for a gene to function as an evolutionary fulcrum, one at which a single mutation alone has a strongly advantageous effect on a trait. If a genome contains few genes on which the phenotype depends in an all-or-nothing manner, then waiting for variants to arise in any of these loci could take too long.

Another possibility is that, even if a single mutation could give rise to a new trait, it would be creating a specialist, one that grows well in one environment but poorly in others. Refining several aspects of a trait with individual genetic changes could be a more flexible way to satisfy more of an organism’s constraints.

Also, despite appearances, polygenic evolution might be a fairly straightforward route to a fitness benefit. Imagine that a microbe does not need a new allele in every gene of a pathway at once to achieve a new trait. Changes in a few genes could be enough to survive the first onslaught of a stressor, for instance. Through time, the resistance trait would improve further as more variants are acquired.
Bayesian phylogenetic trees of three proteins, PRT (A), GTPCH (B), and FPGS (C), involved in the folate metabolic pathways in *E. hellem* and *E. romaleae* (shown in white on blue) but absent from other microsporidians. Numbers at nodes representing Bayesian posterior probabilities (left) and bootstrap proportions (right) are indicated when higher than 0.8 and 70%, respectively. The scale bar corresponds to the estimated number of amino acid substitutions per site. All trees are shown unrooted. (Adapted with permission from Jean-François Pombert et al., Proc. Natl. Acad. Sci. USA 109:12638–12643, 2012.)
**Strength in Numbers**

Whether and how these arguments are broadly relevant for microbial adaptation remains to be seen. Empirically, there is no doubt that multiple loci can control trait differences between microbes. For example, recent studies pinpointed loci that control polygenic drug resistance in the malaria agent *Plasmodium falciparum* and polygenic virulence in the tree-rot fungus *Heterobasidion annosum*.

If natural selection acts on many phenotypes like these in the wild, then the causative loci will show striking polygenic signals in molecular data and will be ripe for discovery and validation. Further, the more genes underlie a trait, the more strongly statistical measures will support an inference of non-neutral change.

Tests for polygenic evolution can spot signals in groups of genes that are defined based on shared annotation, co-regulation, or known interactions. A geneticist may not know the right group to examine but, instead, could screen many groups in a first pass, focusing on the top scorers for biological follow-up.

For this reason, a search for evidence of polygenic evolution needs nothing more than a genome or transcriptome, and has just as much potential to land on an interesting evolutionary story as does a classic test of one gene at a time. Simple and polygenic molecular-evolution methods have their strengths, and each complements the other. Together, these strategies are helping to realize the promise of genomics—the ability to predict, based on data from one experiment, which traits matter to an organism and how they arose.

Rachel B. Brem is Assistant Professor of Genetics, Genomics and Development in the Department of Molecular and Cell Biology, University of California, Berkeley.

**Suggested Reading**


...