Understanding Complex Genetic Systems

Laura F. Landweber

I study the origin of novel genetic systems in microbial eukaryotes, bringing a strongly mechanistic approach to understanding evolution and genome diversity. My lab has shown that the surprisingly sophisticated variations on DNA and RNA processing in microbial eukaryotes create an imaginative playground for genome architecture and genetic systems. Some of their pathways erode the notions of a gene (e.g. scrambled genes and RNA editing) and even Mendelian inheritance, reminding us that a genome sequence can be a far cry from knowledge of its products. Genome-wide DNA rearrangements occur in diverse organisms, and contribute to many human diseases, but their extreme exaggeration in ciliates, particularly *Oxytricha,* makes it an ideal model system to study epigenetic phenomena. My laboratory is currently focused on the mechanism and evolutionary origin of this remarkable RNA-mediated process in *Oxytricha*.

By combining molecular, evolutionary, theoretical, and synthetic, experimental biology, we recently discovered an RNA-guided epigenetic mechanism underlying programmed genome rearrangements (Nowacki *et al.* 2008 and Fang *et al.* 2012 *Cell*), as well as the transitional, or putative intermediate, forms that this complex system passes through during development (Möllenbeck *et al.,* 2008) and evolution (Chang *et al.* 2005) plus its evolutionary origin from a novel mutualistic relationship with thousands of active transposons (Nowacki *et al*. 2009). Nowacki *et al*. 2008 also unveiled new roles for RNA, a new class of long noncoding RNAs, and a new mechanism for epigenetic wiring (and rewiring) of cellular programs, which may relate to phenomena of genome remodeling that contribute to cancer as well (Fang and Landweber, 2013). We found that the long, noncoding RNAs are also templates for DNA repair (Nowacki *et al*. 2008) and can regulate gene dosage and chromosome copy number (Nowacki *et al*. 2010). We further discovered an extensive cache of small RNAs (piRNAs) that protect DNA regions against loss or cleavage during genome rearrangement (Fang *et al*. 2012). Complementary to this discovery of such tiny guardians of the genome, we found that cytosine methylation of DNA can mark DNA sequences for deletion during *Oxytricha* genome rearrangement (Bracht *et al*. 2012).

Our previous work also made fundamental contributions to other areas of research, including creating the first "RNA computer" (Faulhammer *et al.* 1999), which held the record for molecular computation for several years, as well as visualizing natural biological systems as computation (e.g. Landweber and Winfree 2002; Landweber and Kari 1999) and turning the study of the origin of the genetic code into a quantitative, rigorous testing ground (e.g. Knight and Landweber 1998, 2000; Landweber 1999; Freeland *et al*. 2000; Ronneberg *et al.* 2000, 2001; Knight *et al*. 2001 *Nat Rev Genet),* firmly disproving that it was ever a frozen accident. Because ciliates bear the richest diversity of alternative genetic codes, they even helped us illuminate the rules underlying modern transformation of the genetic code itself (Lozupone *et al.* 2001; Liang *et al*. 2005 *JME*). Several purely bioinformatic studies have examined different aspects of ancient and modern protein translation using comparative genomics (e.g. Knight *et al*. 2001, Liang *et al*. 2005, both *Genome Biology*) and structural data (Freeland *et al*. 1999; Liang & Landweber 2005, Liang *et al.* 2005, *RNA*), while Liang & Landweber (2006, 2008, 2013) illuminated hundreds of dual-coding regions in the human genome. We also continue to investigate pathways for early evolution from an RNA world. Both experimental (Landweber and Pokrovskaya 1999; Vlassov *et al*. 2004) and theoretical or computational studies (e.g. Vlassov *et al*. 2005; Goldman *et al*. 2013) have addressed specific issues relating to the origin of coding information, including the transition from RNA- to DNA-based genomes (Goldman and Landweber, 2012 and submitted).

The Mechanism of Genome Unscrambling

A key advancement in my lab has been a close examination of the molecular basis for programmed DNA rearrangements in ciliates. During differentiation of its somatic macronucleus, *Oxytricha trifallax* destroys roughly 95% of its germ-line, severely fragmenting its chromosomes, and then unscrambles hundreds of thousands of remaining fragments by translocation or inversion. Programmed DNA deletion, inversion, and permutation events provide the main experimental system we are using to explore the mechanism underlying complex genome rearrangements. Massive DNA rearrangements reconstruct a set of ~16,000 gene-sized 'nanochromosomes' (Swart *et al*., 2013) for expression in *Oxytricha*'s somatic macronucleus. These DNA acrobatics include programmed loss of intergenic DNA as well as internal eliminated sequences within genes, accompanied by assembly and reordering of coding segments from precursor germline segments sometimes present in a scrambled, permuted order. Short sequence repeats at the junctions between coding and noncoding sequences facilitate reconstruction of intact chromosomes during development.

The molecular mechanisms of genome unscrambling are beginning to come to light. Recent studies in the distantly related ciliates *Tetrahymena* and *Paramecium* suggested that small RNAs might function to specify eliminated sequences by a mechanism similar to RNA-mediated gene silencing (reviewed in Nowacki *et al*. 2011 and Bracht *et al*. 2013). We recently demonstrated an epigenetic mechanism for DNA unscrambling (Nowacki *et al.* 2008). Our model proposes the existence of long, noncoding RNA templates, which contain complete copies of the previous generation's reorganized genome, and that these molecules guide assembly of cryptic germline fragments in a correct order and orientation (Angeleska *et al.* 2007). Nowacki *et al.* 2008 offered several lines of evidence for the transient existence and action of these long, RNA templates. We successfully disrupted gene unscrambling in *Oxytricha* by either microinjecting foreign nucleic acids or feeding the cells bacteria expressing specific double-stranded RNA to target the templates. These experiments resulted in aberrant gene unscrambling in the progeny, either blocking the process completely or producing incorrect DNA patterns. This exciting line of experiments provided us with the first opportunity to parse the specific steps in this complex biological process and permitted us to develop *Oxytricha* as a powerful model system. Microinjection of synthetic DNA or RNA templates demonstrated that these artificial molecules can orchestrate new genome rearrangements in *Oxytricha*, reprogramming the DNA rearrangement pathway for multiple sexual generations (Nowacki *et al*. 2008). This strongly suggests that the long RNA molecules guide genome rearrangement, supporting our epigenetic model. More broadly, these experiments revealed a new class of long, noncoding RNAs, a new role for RNA in biology, and a novel RNA-guided pathway to regulate genome architecture, as well as a specific mechanism for the transmission of an acquired genetic state.

We are also beginning to study the biochemical mechanism of DNA rearrangement, which we found is dependent on thousands of germline-encoded transposases, based on genome-wide disruption of DNA rearrangement after simultaneous RNAi knockdown of three candidate transposase families present in the germ-line (Nowacki *et al*. 2009). Currently, armed with RNA-seq data during the time-course of genome rearrangement (Swart *et al*. 2013), we are testing several more candidate genes via knockdown for potential roles in this process. For example, Fang *et al*. (2012) found that a Piwi protein, Otiwi1, is required for the viability of sexual progeny and for the accumulation of 27 nt Piwi-interacting small RNAs (piRNAs) that derive from the maternal nucleus and specify genome regions for retention. In a key experiment, injecting synthetic piRNAs that correspond to normally-deleted regions leads to their retention in subsequent generations (Fang *et al*., 2012). Thus, these findings also highlight small RNAs as

powerful transgenerational carriers of epigenetic information for genome remodeling. Together, these experiments showcase *Oxytricha* as a model for RNA-mediated inheritance.

The Origin and Evolution of Scrambled Genes and Genome Architectures

Earlier work proposed and tested two different step-wise models for how genes can become extensively fragmented and scrambled over evolutionary time (Chang *et al*. 2005; Wong and Landweber 2006). The data support a model of frequent germline recombination between coding and noncoding DNA, and this trend leads to exaggerated patterns and increased levels of genome scrambling and complexity over time (Chang *et al.* 2005). In a related study, we found that the distinction between retained and deleted sequences in subtelomeric regions can even be switched during recent, within-species evolution (Möllenbeck *et al.* 2006), possibly mediated by sequence evolution in piRNAs and their coverage or abundance, one of many directions we are beginning to explore. Two essays in *Science* (Landweber 2007, 2008) discussed the evolutionary origin of "genomes in pieces" and whether the presence of such a shattered genome architecture might actually confer a selective benefit to some (mostly asexual) organisms. While one might not expect such a Rube Goldberg genetic architecture as *Oxytricha*'s to confer an adaptive benefit to the host organism, there may be hidden charms. Firstly, in *Oxytricha*'s genetic system, there are two vehicles for inheritance: RNA and DNA. Rather than just the usual storage of genetic information in DNA, *Oxytricha* can also pass epigenetic information from parent to offspring in the form of RNA (Nowacki *et al*. 2008; Fang *et al*. 2012) and we are currently sequencing the complete RNA cached copy of its genome during development. While the conjecture that this mechanism confers a selective advantage has not been tested in the laboratory, it would be amenable to long-term evolution experiments.

Previous work discovered and characterized new scrambled genes and their orthologs in various species on a case by case basis (e.g. Kuo *et al.* 2006; Chang *et al.* 2005*,* 2006). The current germline genome project (see below) and our pipeline for analysis offers an unprecedented opportunity to decipher the global patterns and process of rearrangement of the set of over 3,500 scrambled genes that we recently discovered in *Oxytricha*. Comparative genomics is clearly the next step, extending these tools to both closely- and distantly-related species, so that we can examine how a lineage's history allowed it to acquire such a chaotic genome architecture. Comparative genomics over the time-course of development, itself, is also underway, to permit us to dissect the cascade of events that produce a new, rearranged genome.

One Cell: Two Complex Genomes

Sequencing the genomes for both the somatic macronucleus (Swart *et al*. 2013) and the germline micronucleus (Chen *et al*., submitted) of *O. trifallax* (haploid genome sizes ~ 50 Mb and 0.5 Gb, respectively) has provided tremendous insight into the limits of genome architecture. While the macronuclear genome project (which one anonymous reviewer described as "an epic genome project, an important landmark") employed three sequencing platforms (Sanger, 454, and Illumina, requiring us to develop a novel computational pipeline to co-assemble the mixture of three types of data; Swart *et al*. 2013) and NHGRI, NIGMS, and NSF support, the germline genome project has all been done at Princeton with Illumina data and my R01 and NSF support.

Oxytricha's macronucleus has a unique genome architecture (Swart *et al.* 2013; cover of *PLoS Biology*). The fully-sequenced genome contains over 16,000 tiny chromosomes averaging just 3 kb. Most encode single genes, amplified to a few thousand copies. The smallest chromosome is 469 bp and encodes a 98 aa protein, whereas the largest chromosome is 66 kb and encodes a single, enormous *Titin-like* protein. Alternative fragmentation of multi-gene

chromosomes (containing up to 8 genes) produces isoforms with shared sequence and high levels of variation. Another remarkable feature of *Oxytricha*'s macronuclear genome is its inordinate fondness for telomeres, with over 60 million chromosome ends per macronucleus!

The sequencing and assembly of *Oxytricha*'s massively scrambled micronuclear germline genome has required a tour de force effort of experimental and computational biologists in my group. No other sequenced genome bears the complexity of *Oxytricha*, which has over 3,500 scrambled genes, thousands of which overlap, and constructs a somatic genome from an obscure set of ~250,000 tiny DNA pieces (242,921 is our precise estimate at present) scattered over long stretches of the germline archive. These gene pieces assemble to produce a new somatic genome during every round of sexual reproduction and nuclear development.

Efforts in the coming years will be on both comparative genomics and on building and testing a model that puts together our observations of the roles of small and long, noncoding RNAs, together with DNA and histone modifications and studies of chromatin, to elucidate the cascade of events that take place during genome rearrangement. The field of epigenetics, particularly RNA-mediated epigenetic inheritance, is presently taking off, and our discoveries have been on the forefront of this discipline, revealing both new classes and functions of small and long, noncoding RNAs, and new roles for RNA in regulating a range of phenomena, from genome rearrangement to RNA-mediated DNA repair and gene dosage.

The genomic and functional resources available in my lab will pave the way for future studies of other species with extensive genome rearrangements, including non-ciliate microbial eukaryotes (Landweber, 2007). I have had a successful tradition of working on a range of non-canonical model systems, including kinetoplastids (Landweber and Gilbert 2003, 2004) and slime molds (Horton and Landweber, 2000, 2002) and occasionally more conventional models (Liang *et al.* 2005 *Genome Biol*; Liang & Landweber, 2006, 2007; Cavalcanti *et al.* 2006; Fang *et al*. 2012 *Biol Dir*). For several years I co-chaired the NHGRI Working Group on Comparative Genome Evolution, which selected new organisms for genome sequencing and entered many projects into the pipeline, and we are currently collaborating on the *Physarum* genome assembly. A proposed new direction would be to study the maximally fragmented mitochondrial genomes of diplonemids (Landweber 2007) to decipher the functional events of RNA *trans*-splicing that reassemble its mitochondrial transcripts from hundreds of gene pieces, each encoded on a separate DNA molecule. It is plausible that more roles for noncoding RNAs in regulation will be uncovered in this system. We are also examining the possible associations of RNA-mediated DNA rearrangements (Nowacki *et al*. 2008) with disease, as well (Fang *et al*. 2012 *Biol Direct*).

Oxytricha research in my lab continues to expand in several directions. Single-cell experiments are allowing us to explore cytoplasmic dominance of RNA-mediated epigenetic effects. Mass spectrometry facilitates the discovery of proteins present in differentiating nuclei or that interact with key players during development (Fang *et al*. 2012). So far, this has confirmed expression of dozens of germline-limited genes that we discovered from the genome sequence (Chen *et al*. submitted). With their expression limited to development, many of these proteins are candidates for roles in genome remodeling, which we can test by knockdown. In addition, we continue to develop tools to improve the ease of using *Oxytricha* as a model system, and we are currently exploiting piRNA-programmed sequence retention for gene knockout.

The combination of high resolution genomic analyses and functional molecular experiments in *Oxytricha* and other species will lead to a broader understanding of the acrobatic processes of genome rearrangement that can accumulate in an organism over evolutionary time. On a grand scale, these events at the genomic level contribute to the diversification of life on our planet, showcasing the astonishing range of genetic mechanisms present in microbial eukaryotes.