Eliminating the Cost of Sex: Asexual Clonal Lineages Amidst Sexual Eukaryotic Microbes

Walter L. (Larry) Ruzzo
University of Washington, USA

Joint work with T Chiang, J Koester, C Berthiaume, N Hiranuma, M Parker, V Iverson, R Morales, A Sarwate & E Armbrust
Quick review: genetics of sex

• Mitosis
  • all chromosomes duplicated
  • one division of a (typically) diploid cell
  • carefully allocate Chr's to the two (diploid) daughter cells

• Meiosis & Recombination
  • all chromosomes duplicated
  • homologous chromosomes pair (pairs of sister chromed pairs)
  • recombination occurs between homologous pairs
  • then cells divide, twice, yielding 4 haploid gametes, each with chromosomes that are a mix of maternal/paternal
Hardy-Weinberg Equilibrium

Very simple model of distribution of alleles in a population, assuming:

- sexually reproducing diploid organisms
- non-overlapping generations
- random mating
- infinite population size
- equal allele frequencies in both sexes
- no migration, mutation or selection
Hardy-Weinberg Equilibrium

• Suppose 2 alleles, say A & a, exist at one site with population-wide frequencies of p and q = 1 - p

• What are frequencies of the diploids AA, Aa, aa?
  • AA: \( p^2 \)
  • Aa: \( 2pq \)
  • aa: \( q^2 \)

• NB: \( p^2 + 2pq + q^2 = (p+q)^2 = 1 \)

• And assume no linkage, so adjacent site B & b, etc. will independently appear with analogous probabilities \( \tilde{p}^2, 2\tilde{p}\tilde{q}, \tilde{q}^2 \)
Sex is ancient, ubiquitous

+ Allows deleterious alleles to be shed (or not)
+ Accelerates mixing of alleles (for better or worse)

- BUT: Meiosis is complex, slow & expensive, Finding mates is hard, Only 1/2 of genes are passed on, …

- Ancient eukaryotic asexuals are rare, but theoretical support for benefit of sex is still debated
Diatoms

http://sarahjanemaki.com/diatoms-3/
Diatoms

• First formally described in scientific literature by Danish naturalist Otto Friedrich Müller, 1783.

• Photosynthetic, unicellular, mostly aquatic, eukaryotes

• Plausibly about the same age as land plants, but at least an order of magnitude more species-rich

• Also noted for high in-species genetic diversity

• Estimated to contribute 20–40% of primary production

• Silica cell wall; they dominate oceanic SiO$_2$ cycling
Phytoplankton bloom across the Barents Sea off the coast of Cape Nordkinn in Norway.

http://www.esa.int/Our_Activities/Observing_the_Earth/Space_for_our_climate/Earth_from_Space_Summer_in_bloom
Sex and the Diatom

- Eukaryotes, normally diploid
- Most cell divisions are *mitotic*
  - 2 diploid daughter cells, each gets one valve (1/2 of cell wall)
  - New valve *inside* old ⇒ one daughter is smaller
- Occasionally undergo *meiosis*
  - Haploid gametes, die if they don’t fuse with a partner (unlike yeast, e.g.)
- Triggers for sex are largely unknown; one is thought to be cell-size reduction – Auxospore (fertilized egg) outgrows its valve & makes new, larger ones
Thalassiosira pseudonana

Photo: N. Kröger
Thalassiosira pseudonana

- A marine diatom
- Named long ago
- In continuous culture since 1958
- “Cosmopolitan” = found all over the world
- First diatom genome sequence (Sanger-based, 2004)
  - Diploid, ~ 32 Megabases, “SNP” every 100-200bp
Re-Sequenced 7 Isolates

Biogeography – correlate diversity with geography
Re-Sequenced 7 Isolates

- Goal:
  - Biogeography – correlate diversity with geography
- Findings:
  - There is (almost) no geographical diversity!
  - 5 of 7 are nearly identical, genetically
  - Down to the level of sharing heterozygous positions
  - Why? They are obligate asexual clones!
  - And they rapidly colonized the world’s oceans
  - The other 2: we see nothing that contradicts sex and Hardy-Weinberg (tho sex has never been observed)
Non-reference Read to Coverage \(\approx\) Alternate Alleles

\[
R = \frac{\text{nonreference}}{\text{coverage}}
\]

**Reference**

```
CTGCTGGTAAATAGGAAGGCTA
```

**Aligned Reads**

```
C TG C T G G T A A T A G G A A G G C T A T
C TG C T G G T A A T A G G A A G G C T A T
C TG C T G G T A A T A G G A A G G C T A T
C TG C T G G T A A T A G G A A G G C T A T
C TG C T G G T A A T A G G A A G G C T A T
G T G C T G G T A A T A G G A A G G C T A T
G T G C T G G T A A T A G G A A G G C T A T
G T G C T G G T A A T A G G A A G G C T A T
G T G C T G G T A A T A G G A A G G C T A T
G T G C T G G T A A T A G G A A G G C T A T
```

**R = 1/2**
Non-reference Read to Coverage is Imperfect Proxy

\[ R = \frac{\text{nonreference}}{\text{coverage}} \]

\[ \text{Aligned Reads} \]

\[ \text{Reference} \]

\[ R = \frac{3}{10} = 30\% \quad \text{\sim} 0 \quad 9/10 \]
Allelic-Distribution of Two Isolates

Wales, UK - CCMP1013
Venice, Italy - CCMP3367

Blue: Histogram of R-value distribution (Chr 1)

R ≈ 0 : homozygous, reference
R ≈ 1 : homozygous, non-reference
R ≈ 0.5: heterozygous (ref + non-ref)

Why ≠ 0/1?
Seq & map errs

Why ≠ .5?
Sampling
Matches H-W Expectation

\[
\begin{align*}
\frac{\text{Heterozygous}}{\text{Homozygous Alt}} & \approx \frac{2}{1} \\
\end{align*}
\]

\( \text{blue} = \text{Wales}; \quad \text{orange} = \text{simple theoretical H-W model} \)
A Digression: Mapping Bias

Coverage around SNPs (Chr1) CCMP1335 (NY)

Coverage

N = 15582
○ Coverage

Average Coverage

Relative Position from a SNP

-100 -50 0 50 100
New York does not match HWE

- Missing homozygous non-reference peak at R = 1
- NY is the reference strain
New York does not match HWE...nor does Washington

New York, USA - CCMP1335

- Missing homozygous non-reference peak at $R = 1$
- NY is the reference strain

Washington, USA - CCMP1015

- WA is not!
Not only the same distribution but the same heterozygous positions

- 10K randomly selected genomic positions
- Plotted the R-value of NY against WA
- Concentration of sites are in two locations: (0,0) & ($\frac{1}{2}$, $\frac{1}{2}$)
- There are no points at (0,1) & (1,0)
Strong agreement of heterozygous positions

Heterozygous concordance with the reference is at least 96%.
Heterozygous concordance implies extreme departure from HWE

We have detected:

1. Clonal cultures from 5 dispersed regions
2. \( \geq 96\% \) concordance in SNPs

What is the probability that this population is in Hardy-Weinberg Equilibrium along with the above constraints?

**Simple Binomial Model**

\[ p \leq 1.2 \times 10^{-29} \]
Data Summary

• 5 of 7 isolates share 96% or their SNPs

• Essentially no mixing of alleles as expected under Hardy-Weinberg, e.g. heterozygous positions are abundant in all individuals but (almost) never re-assort, (almost) never result in homozygous but non-reference positions, ...

• Estimated crossover rate in the 5 is ~20x lower than in the other 2

• CONCLUSION: they are obligate asexuals, reproducing exclusively by mitotic cell division for ~1000 years
Spacial uniformity of heterozygous sites reveals evolutionary history

Isolates from 5 Dispersed Ecosystems Share Loss of Heterozygous Regions

The blue regions are nearly homozygous regions (aka SNP deserts)
LoH Events Happened Nearly Simultaneously

B. CCMP 1335 Large Deserts vs Intervening Regions
T. Pseudonana History?

Shared SNPs:
- Venice, Italy (3367)
- Wales, UK (1013)
- Virginia, USA (1007)
- Perth, W. Australia (1012)
- Washington, USA (1015)
- N. Pacific Gyre (1014)

Inbreeding LoH/LoS:
- a: only in 1014 - 559
- b: only in 1335 - 620
- c: only in 1015 - 2070
- d: only in 1012 - 611
- e: only in 1007 - 321
*: in bcde - 9777
  in bc - 410
  in de - 1150
Discordant SNPs - 81744

Number of Shared SNPs:
- 38232 in 5
- 67223 in 7
- 87584 in 2
- 93481 only in 1013
- 84335 only in 3367
Summary

- A population bottleneck / inbreeding caused a loss of heterozygosity in a sub-population of T. pseudonana
- A functional loss of sex occurred within this sub-population
- This obligate asexual lineage spread across the world's oceans
- At least 5 CCMP isolates are descendants of this lineage
- At least 2 CCMP isolates maintain sexual reproduction in the wild
Re-Sequenced 7 Isolates

Clonal/Asexual:
Global Invasion
Implications

• A clonal global dispersal implies the existence of a general purpose genotype for *T. pseudonana*.

• It is unlikely that this type of obligate asexuality is unique to a single marine microbe.

• Environmental isolates may be biased in favor of such genotypes explaining the lack of sexual reproduction in culture despite attempted induction.

• This clonal sub-population is susceptible to global disruption by disease or environmental perturbations.

• Counters the classical assumption - genetic diversity correlates with geographic distance (biogeography).
Acknowledgements

• Collaborators:
T Chiang, J Koester, C Berthiaume, N Hiranuma, M Parker, V Iverson, R Morales, A Sarwate & E Armbrust

• Funding:
Systematic DNA Loss in Culture
(⇒ No Contamination)

<table>
<thead>
<tr>
<th>Isolation Date (Year)</th>
<th>Hemizygous Deletion (Kilobases)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>8.2</td>
<td>New York, USA</td>
</tr>
<tr>
<td>1964</td>
<td>26.7</td>
<td>Virginia, USA</td>
</tr>
<tr>
<td>1965</td>
<td>21.4</td>
<td>Perth, W. Australia</td>
</tr>
<tr>
<td>1971</td>
<td>4.5</td>
<td>NPG</td>
</tr>
<tr>
<td>1973</td>
<td>9.7</td>
<td>Wales, UK</td>
</tr>
<tr>
<td>1985</td>
<td>11.7</td>
<td>Washington, USA</td>
</tr>
<tr>
<td>2006</td>
<td>9.0</td>
<td>Venice, Italy</td>
</tr>
</tbody>
</table>