Signatures of a population bottleneck can be localised along a recombining chromosome

Céline Becquet, Peter Andolfatto
Bioinformatics and Modelling, INSA of Lyon
Institute for Cell, Animal and Population Biology, Edinburgh

celine_becquet@netcourrier.com

June, 2003
Introduction

Distinguishing between Bottleneck and Hitchhiking

- Natural genetic variability
  - *Neutral Theory (Kimura 1968, 1983)*
    - Neutral mutations, Recombination, Genetic drift
  - *Natural selection affects variability*
    - Localised effects
    - Functional genomic regions can be detected
  - *Demography also affects variability*
    - The entire genome is affected in a similar manner
Introduction

Tests of neutrality

- Detect Departure from the Standard neutral model (SNM)
  - *Neutral mutation*
  - *Sampling from panmictic population*
  - *Constant population size*

- Measure variability on
  - *Single locus*
  - *Multi-locus*

- Demography affects these tests as well
  - *But genome wide vs. localised effects*
  - Enable detection of departure from the SNM
Introduction

Data from Drosophila and Humans

Show complicated demographic history

- Origin in Africa
- Differentiation between non-African and African pops.
  - Linkage disequilibrium (LD) higher in non-African pops.

- Suggest an “out of Africa” bottleneck
Introduction

**Our work: detecting bottleneck’s signature**

- Simulate gene genealogies
- Under various models of bottlenecks
  - Based on “best guess” parameters for both Drosophila and humans
- Observe
  - Behaviour of some test of the SNM
  - Patterns of variability across the recombining chromosome

  How often a bottleneck will lead us to the incorrect conclusion that selection has occurred in a genomic region of interest?
Material & methods

Bottleneck Simulations

Coalescent simulations

- Hudson (2002) Program
  - Standard Neutral Model (SNM)
  - Simulations based on
    - $\theta = 4N\mu L$
    - $S$: number of segregating sites
  - Finite-sites recombination model
    - $\rho = 4NrL$
  - Parameters of bottleneck
    - Number of intervals of population size changes
    - Reduced population size ($N_b$)
    - Starting time ($T_b$)
    - Duration of the bottleneck ($T$)
Simple Step Bottleneck

Population Size

Time

N0

Nb

Tb

T
Material & methods

Bottleneck Simulations

- Drosophila populations
  - \( N_0 = 5,000,000 \) individuals
  - Average variability reduced by 15% or 50%
  - Population mutation rate
    - Single locus: \( \theta = 15 \) (500 recombining bp)
    - Large survey region
      - \( \theta = 1200 \) over 40,000 recombining bp
      - Windows of 500 bp, 50 bp step
  - Recombination rate: \( \rho = 3\theta \) or \( 15\theta \)
  - Sample size of 15 chromosomes
Material & methods

Bottleneck Simulations

- Human populations
  - $N_0 = 12,000$ individuals
  - Variability reduced by 35%
  - Population mutation rate
    - Single locus tests: $\theta = 3$ over 2500 recombining bp
  - $\rho = \theta$
  - Sample size of 15 chromosomes
Material & methods

Statistical tests

- **Level of polymorphism** *(Kreitman and Hudson’s 1991)*
  - Number of segregating sites \((S)\)
  - Low level of diversity expected under a **Bottleneck**
  - One-tailed test at 5%
Material & methods

Statistical tests

- Frequency spectrum

  - *Tajima’s (1989) D*
    - Normalised difference between $\theta_W$ (Watterson 1975) and $\pi$ (Tajima 1983)
    - $E(D)>0$, $<0$ or $=0$ under a bottleneck
    - One-tailed tests at 5% for +ve and -ve values

  - *Fay & Wu’s (2000) H*
    - $H = \pi - \theta_H$
    - -ve under selection
    - Can be affected by demography too
    - One-tailed test at 5%
Material & methods

Statistical tests

- **Linkage disequilibrium**
  - *Number of distinct haplotypes (K Strobeck 1987)*
    - Recent bottleneck or periodic bottlenecks $\rightarrow$ small K
    - One-tailed test at 5%
  - *Frequency of the most frequent haplotype ($f_{MFH}$)*
      - Recent bottleneck $\rightarrow$ strong haplotype structure
      - One-tailed test at 5% on upper values
Material & methods

Statistical tests

- Statistical tests performed
  - Considering 15 chromosomes
  - 10,000 repetitions
  - Assuming no intragenic recombination

  So, statistics are conservative
# Results: Single locus rejection probability

## Step-recovery models for Drosophila

### Rejection probability for single loci under step-recovery bottleneck models for *Drosophila* \( (\rho = 3\theta) \).

<table>
<thead>
<tr>
<th>Bottleneck</th>
<th>( \theta/\theta_0 )</th>
<th>( S ) (( \mu ), reject)</th>
<th>( K ) (( \mu ), reject)</th>
<th>( f_{MFH} ) (( \mu ), reject)</th>
<th>( D ) (( \mu ), reject 5%, 95%)</th>
<th>( H ) (( \mu ), reject)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mild</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent ( T_b = 2000 \text{ ga} )</td>
<td>0.82</td>
<td>40.1, 0.005</td>
<td>8.2, <strong>0.15</strong></td>
<td>0.28, <strong>0.07</strong></td>
<td>0.41, 0.001, 0.03</td>
<td>-1.28, 0.02</td>
</tr>
<tr>
<td>Old ( T_b = 120,000 \text{ ga} )</td>
<td>0.83</td>
<td>40.7, 0.005</td>
<td>10.1, 0.02</td>
<td>0.23, 0.03</td>
<td>0.36, 0.001, 0.03</td>
<td>-1.19, 0.02</td>
</tr>
<tr>
<td>Eqb. Pop.</td>
<td>0.81</td>
<td>39.3, 0.002</td>
<td>12.6, (&lt;0.0001)</td>
<td>0.16, 0.001</td>
<td>-0.01, 0.001, 0.001</td>
<td>0.00, 0.01</td>
</tr>
<tr>
<td><strong>Severe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent ( T_b = 2000 \text{ ga} )</td>
<td>0.50</td>
<td>24.3, <strong>0.07</strong></td>
<td>3.5, <strong>0.96</strong></td>
<td>0.55, <strong>0.56</strong></td>
<td>0.88, 0.04, <strong>0.35</strong></td>
<td>-3.94, <strong>0.21</strong></td>
</tr>
<tr>
<td>Old ( T_b = 120,000 \text{ ga} )</td>
<td>0.52</td>
<td>25.4, <strong>0.06</strong></td>
<td>5.9, <strong>0.39</strong></td>
<td>0.47, <strong>0.36</strong></td>
<td>0.69, 0.04, <strong>0.27</strong></td>
<td>-3.77, <strong>0.19</strong></td>
</tr>
<tr>
<td>Eqb. Pop.</td>
<td>0.50</td>
<td>24.5, 0.01</td>
<td>11.2, 0.0002</td>
<td>0.20, 0.002</td>
<td>-0.02, 0.01, 0.01</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td>Ancestral</td>
<td>1.00</td>
<td>48.8, -</td>
<td>13.1, (&lt;0.0001)</td>
<td>0.14, 0.0006</td>
<td>-0.01, 0.0003, 0.0008</td>
<td>0.05, 0.004</td>
</tr>
</tbody>
</table>

Simulations based on observed variability from the simulations with bottleneck.

*Dispersal from Africa: 10-15 kya ~ 120,000 ga*

*Founding of North America pops.: < 400 ya ~ 2,000 ga*
**Results: Single locus rejection prob.**

*Step-recovery and single step models for Human*

Rejection probability for single loci under different bottleneck models for Humans

<table>
<thead>
<tr>
<th>Bottleneck</th>
<th>(\theta/\theta_0)</th>
<th>(S) ((\mu), reject)</th>
<th>(K) ((\mu), reject)</th>
<th>(f_{MFH}) ((\mu), reject)</th>
<th>(D) ((\mu), reject 5%, 95%)</th>
<th>(H) ((\mu), reject)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recent (T_b = 600) ga</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step-Recovery</td>
<td>0.65</td>
<td>6.2, <strong>0.09</strong></td>
<td>3.8, <strong>0.10</strong></td>
<td>0.55, <strong>0.08</strong></td>
<td>0.42, 0.03, <strong>0.16</strong></td>
<td>-0.58, <strong>0.12</strong></td>
</tr>
<tr>
<td>Simple step</td>
<td>0.65</td>
<td>6.1, <strong>0.09</strong></td>
<td>3.6, <strong>0.12</strong></td>
<td>0.55, <strong>0.08</strong></td>
<td>0.56, 0.03, <strong>0.19</strong></td>
<td>-0.57, <strong>0.12</strong></td>
</tr>
<tr>
<td><strong>Old (T_b = 2500) ga</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step-Recovery</td>
<td>0.64</td>
<td>6.2, <strong>0.08</strong></td>
<td>4.3, <strong>0.05</strong></td>
<td>0.58, <strong>0.07</strong></td>
<td>-0.02, <strong>0.06</strong>, <strong>0.09</strong></td>
<td>-0.55, <strong>0.11</strong></td>
</tr>
<tr>
<td>Simple Step</td>
<td>0.65</td>
<td>6.2, <strong>0.08</strong></td>
<td>3.9, <strong>0.09</strong></td>
<td>0.55, <strong>0.07</strong></td>
<td>0.40, 0.03, <strong>0.15</strong></td>
<td>-0.58, <strong>0.11</strong></td>
</tr>
<tr>
<td>Eqb. Pop.(^c)</td>
<td>0.65</td>
<td>6.3, 0.04</td>
<td>5.1, 0.01</td>
<td>0.49, 0.02</td>
<td>-0.05, 0.03, 0.04</td>
<td>0.02, 0.05</td>
</tr>
<tr>
<td>Ancestral(^d)</td>
<td>1.00</td>
<td>9.8, -</td>
<td>6.5, 0.01</td>
<td>0.39, 0.02</td>
<td>-0.04, 0.03, 0.04</td>
<td>0.01, 0.04</td>
</tr>
</tbody>
</table>

Between emergence of modern humans 2500 ga ~ 50 kya
And introduction of agriculture 600 ga ~12,000 ya
Results: Large survey region

Pattern of variability along a recombining seq.

*Drosophila*

Severe & old Bottleneck $\rho=3\theta$

Simulation with bottleneck

Simulation with equilibrium population
Results: Large survey region

Pattern of variability along a recombining seq.

_Drosophila_

Severe & old Bottleneck $\rho=3\theta$
Results: Large survey region

Pattern of variability along a recombining seq.

Drosophila
Severe & old Bottleneck $\rho=3\theta$
Results: Large survey region

Patterns of variability show

- Bottleneck increases
  - Variance of the statistics (except haplotype tests)
  - Physical scale of the LD
    - The extent to which genealogies are shared
Probability distribution of the size of the rejecting regions

*Drosophila* Severe & old Bottleneck

\[ \rho = 3\theta \]

Simulation with equilibrium population

Simulation with bottleneck
Discussion

Interpretations of Significant tests

- Under a bottleneck, we observe
  - +ve D, -ve H; increased LD

- Data show
  - In Drosophila
    - D is not generally +ve
      - Our model may not fit reality
    - More -ve D in African pops.
      - Growth in the ancestor covers up evidence for bottleneck?
  - In Human
    - D not generally +ve
      - Inadequate sampling, pop growth, or pop structure?
    - More -ve D in African pops.
Discussion

Interpretations of Significant tests

- Comparison between African and non-African often show many significant haplotypes and Fay & Wu’s $H$ tests
  - In Drosophila species (numerous examples)
  - In Human (OR, Gilad and Lancet 2003 and others)

- But our results suggest that
  - $H$ and haplotype tests not conservative under pop. size changes
  - These observations could be interpreted as bottleneck signature
Discussion

Multi-locus analysis not optimal

- Heterogeneity argument used to argue for selection
  - In Droso. (Mousset et al. 2003, Andolfatto et al. 1999, Hudson et al. 1997, and others)
  - In Human (Kayser et al. 2003, and others)

- We show
  - Bottleneck increases
    - Variance of most statistics
    - Physical scale of LD across a chromosome
  - Observed heterogeneity in polymorphism patterns may be consistent with demography
**Conclusion & Prospects**

- **Demographic history is important**
  - Changes in population size may be common in the history of most species
  - Other models of demographic changes should also be studied

- **Finding evidence for selection may be difficult**
  - Need specific methods to distinguish demography from selection
**Conclusion**

& **Prospects**

- **Combining tests together?**
  - *The studied tests are too correlated*

- **Explicitly model bottleneck?**
  - *But unknown parameters*

- **Lazzaro and Clark (2003)***
Thanks

ICAPB, Edinburgh  INSA, Lyon

Nick Barton  Guillaume Beslon

Peter Andolfatto  Jean-Michel Fayard

Nick Barton’s lab  Hedi Soula

And all the 1rst floor
Working atmosphere?

Who hell is this?

Ugly boy

Sweet lips

The kid

Mouth from the south