Fitness constraints on horizontal gene transfer

Dan I Andersson
University of Uppsala,
Department of Medical Biochemistry and Microbiology,
Uppsala, Sweden

GMM 3, 30 Aug--2 Sep, Oslo, Norway
Acknowledgements:

Uppsala University
Peter Lind

Collaborators
Otto Berg
Chuck Kurland
Bacterial genome evolution

- Gene acquisition
- Duplication
- Selection Drift
- Loss of fragments encoding one or more genes
- Inactivation to form a pseudogene
- Erosion by deletional bias

Mira et al. 2001
What genes are transferred and why?

What?
- Common: antibiotic resistance, pathogenicity, metabolic functions, transport: i.e. functions associated with niche-specific adaptations
- Rare: translation, transcription, replication: i.e. niche-independent essential functions

Why?
- Can give immediate benefit
- Complexity hypothesis: proteins with many interactions (structural, regulatory, functional) less likely to be transferred, i.e. a newly introduced orthologue is unlikely to outperform a gene that has co-evolved for a long time with its cognate partners
Potential constraints on HGT

1. Ecological opportunity

Buchnera aphidicola within aphid bacteriocyte

2. Transfer systems

Zaneveld et al. Microbiology 2008

3. Function and expression of transferred gene, i.e. fitness

- Expression signals
- Chromosome structure and sequence
- Silencing

- Structure
- Stability
- Codon usage
- Regulation

- Different environments
- Interactions with other proteins
- Regulation
Questions addressed

I. Examine barriers to HGT in terms of the distribution and magnitude of fitness constraints at RNA and protein level

II. Determine the potential and mechanism of compensatory evolution to ameliorate the fitness costs of HGT
Model system: 70S ribosome

50S large subunit:
23S, 5S, 33 r-proteins

30S small subunit:
16S rRNA, 21 r-proteins
+ several translation factors

Complex structure with high functional and structural integration
Ribosome function directly related to fitness (growth rate)
Ribosomal protein S20 (*rpsT*)

- 86 amino acids
- Binds to 16S rRNA
- Deletion mutants defective in assembly of subunits
- Translational repressor
- Non-essential
Ribosomal protein L1 (rplA)

• 234 amino acids
• Binds to 23S rRNA
• Ejection of deacetylated tRNA
• Deletion mutants defective in translocation
• Translational repressor - L1/L11
• Non-essential
Ribosomal protein L17 \((rplQ)\)

- 127 amino acids
- Binds to 23S rRNA
- Located at exit tunnel opening
- Essential
Experimental design

• Model organism *Salmonella typhimurium* LT2

• Two non-essential and one essential ribosomal protein used as model for orthologous transfer

• Highly sensitive measurements of fitness
Strain construction

- Genes from other species cloned in plasmid used as PCR template
- Genes introduced together with antibiotic marker to replace the native Salmonella gene by Lambda Red recombineering --> regulatory and chromosomal context unaltered
Ribosomal protein genes replaced with orthologues from other species

**Rif operon:**

L1

**Alpha operon:**

L17

**S20 operon:**

S20
## Species comparison of ribosomal proteins

<table>
<thead>
<tr>
<th>Gene/protein species</th>
<th>Nucleotide substitutions</th>
<th>Nucleotide identity (%)</th>
<th>Amino acid substitutions</th>
<th>Amino acid identity (%)</th>
<th>Codon adaptation index</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>rpsT/S20 (Å264 bp/86 aa)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salmonella typhimurium</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.611</td>
<td>46.9</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>3</td>
<td>98.9</td>
<td>2</td>
<td>97.7</td>
<td>0.607</td>
<td>47.1</td>
</tr>
<tr>
<td>Citrobacter freundii</td>
<td>6</td>
<td>97.7</td>
<td>1</td>
<td>98.9</td>
<td>0.608</td>
<td>46.9</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>7</td>
<td>97.3</td>
<td>1</td>
<td>98.9</td>
<td>0.616</td>
<td>47.9</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>34</td>
<td>87.1</td>
<td>14</td>
<td>83.7</td>
<td>0.615</td>
<td>50.6</td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>47</td>
<td>82.2</td>
<td>14</td>
<td>83.5</td>
<td>0.592</td>
<td>39.9</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>68</td>
<td>74.2</td>
<td>23</td>
<td>73.6</td>
<td>0.580</td>
<td>40.3</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>179</td>
<td>32.0</td>
<td>46</td>
<td>47.0</td>
<td>0.561</td>
<td>39.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>rplA/L1 (Å705 bp/234 aa)</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonella typhimurium</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.719</td>
<td>53.3</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>45</td>
<td>93.4</td>
<td>4</td>
<td>98.3</td>
<td>0.688</td>
<td>51.9</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>43</td>
<td>93.9</td>
<td>5</td>
<td>97.9</td>
<td>0.712</td>
<td>52.6</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>71</td>
<td>89.9</td>
<td>14</td>
<td>94.8</td>
<td>0.678</td>
<td>53.0</td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>98</td>
<td>86.0</td>
<td>25</td>
<td>88.0</td>
<td>0.578</td>
<td>43.1</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>190</td>
<td>73.0</td>
<td>37</td>
<td>84.0</td>
<td>0.549</td>
<td>39.0</td>
</tr>
<tr>
<td>Neisseria menigitidis</td>
<td>255</td>
<td>63.9</td>
<td>77</td>
<td>67.1</td>
<td>0.578</td>
<td>43.9</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>313</td>
<td>55.6</td>
<td>117</td>
<td>50.0</td>
<td>0.609</td>
<td>43.6</td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>316</td>
<td>55.0</td>
<td>110</td>
<td>53.0</td>
<td>0.536</td>
<td>36.4</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>439</td>
<td>37.7</td>
<td>185</td>
<td>19.7</td>
<td>0.466</td>
<td>36.6</td>
</tr>
<tr>
<td>Sulfolobus acidocaldarius</td>
<td>418</td>
<td>40.7</td>
<td>179</td>
<td>23.5</td>
<td>0.456</td>
<td>35.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>rplQ/L17 (Å384 bp/128 aa)</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonella typhimurium</td>
<td>8</td>
<td>97.9</td>
<td>1</td>
<td>99.2</td>
<td>0.549</td>
<td>54.6</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>8</td>
<td>97.9</td>
<td>2</td>
<td>98.4</td>
<td>0.559</td>
<td>53.8</td>
</tr>
<tr>
<td>Citrobacter freundii</td>
<td>11</td>
<td>97.2</td>
<td>3</td>
<td>97.7</td>
<td>0.597</td>
<td>55.7</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>26</td>
<td>93.3</td>
<td>8</td>
<td>93.8</td>
<td>0.563</td>
<td>55.3</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>45</td>
<td>88.6</td>
<td>8</td>
<td>93.7</td>
<td>0.475</td>
<td>49.7</td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>95</td>
<td>75.4</td>
<td>15</td>
<td>88.3</td>
<td>0.338</td>
<td>41.2</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>125</td>
<td>66.1</td>
<td>24</td>
<td>80.3</td>
<td>0.341</td>
<td>46.7</td>
</tr>
<tr>
<td>Neisseria menigitidis</td>
<td>198</td>
<td>43.6</td>
<td>65</td>
<td>44.0</td>
<td>0.212</td>
<td>45.7</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>206</td>
<td>44.2</td>
<td>63</td>
<td>48.4</td>
<td>0.340</td>
<td>37.5</td>
</tr>
</tbody>
</table>
Growth rate of homologous L1 mutants

Relative growth rate compared to wild type

Relative growth rate of L1 deletion mutant = 0.3
Growth rate of homologous S20 mutants

Relative growth rate compared to wild type

- LB
- Glucose
- Glycerol
- Acetate

Relative growth rate of S20 deletion mutant = 0.3
Growth rate of homologous L17 mutants

Relative growth rate compared to wild type

- LB
- Glucose
- Glycerol
- Acetate

Relative growth rate of L17 deletion mutant = 0
Competition assays

• Need more sensitive assay than single culture growth rate

• Problems with traditional competitions
  – Many colonies needed for statistical significance
  – Adaptive mutations during experiment (periodic selection)
    ---\( s < 0.01 \text{ not detectable}\)---

  Experimental detection vs. population genetics dilemma

• Solutions
  – Label cells with YFP/CFP and count with flow cytometry
  – Pre-adapt strains for 1000 generations
Competitive fitness of homologous mutants

Grown in LB medium
Conclusions replacement experiments

- Replacement with orthologues from other eubacterial species generally confer small fitness costs (at least from a geneticists view!)

- Differences in amino acid sequence, codon usage and GC content have small effects on function

- Fitness effects still large enough to prevent HGT (at least from a population biologists view!)

  A counter-selected import is essentially unable fix in a bacterial population
  For example, $s = -0.01$, $N_e = N = 10^5$, $P_{fix} = 10^{-436}$
Amelioration of fitness costs associated with HGT

Serial passage of HGT mutants with large fitness costs,
10^6 cells transferred to 1 ml LB --> grow to 10^9 cells--> re-inoculate-->
10 generations of growth per passage

Substantial increases in fitness for all mutants after <300 generations
# Characteristics of compensated mutants

<table>
<thead>
<tr>
<th>HGT donor</th>
<th>Fitness</th>
<th>Generations of growth</th>
<th>Copy #</th>
<th>Size of amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Yeast</td>
<td>0.68--&gt;0.75</td>
<td>250</td>
<td>2-3</td>
<td>44 kbp (rRNA)</td>
</tr>
<tr>
<td>S20 Haemophilus</td>
<td>0.60--&gt;0.93</td>
<td>80</td>
<td>2-4</td>
<td>15-90 kbp</td>
</tr>
<tr>
<td>L17 Helicobacter</td>
<td>0.60--&gt;0.70</td>
<td>200</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

![Diagram showing PCR products and control bands](image)
Compensation of suboptimal functions by gene amplification is common

1. Loss of formyl transferase (fmt gene) compensated by tRNA\textsuperscript{Met} gene amplification (Nilsson et al, PNAS 2006)


4. Low \(\beta\)-lactamase activity compensated by amplification of the \(bla_{TEM-1}\) gene (Sun et al, in preparation)
Why are gene amplifications so commonly found as suppressors?

1. Most sub-optimal functions can be improved by increased dosage of the cognate or non-cognate gene (Patrick WM et al Multicopy suppression underpins metabolic evolvability. Mol Biol Evol. 2007)

2. Gene duplications are several orders of magnitude more common than point mutation ($10^{-5}$ to $10^{-2}$ versus $10^{-10}$)
Connecting HGT and the duplication-divergence model for evolution of new genes

(Hooper and Berg, Genome Biology 2003, “a horizontally transferred bacterial gene has a 2- to 10-fold higher probability of being duplicated as compared to an indigenous gene”)

HGT

Gene amplification

Divergence by mutation

Beneficial but suboptimal function

Increased gene dosage

Two paralogs

Evolved gene
Conclusions:

Surprisingly small fitness effects when replacing phylogenetically distant ribosomal proteins

However, these small fitness effects (s values around -0.01) are still sufficiently large to effectively prevent any HGT in a bacterial population

Partial amelioration of the fitness costs of foreign genes via increased gene dosage (amplification)

HGT might promote evolution of new genes via the duplication-divergence process