Recent Advances Towards An Intraspecific Theory of Human Variation for Digital Models

By Bradly Alicea
freejumper@yahoo.com
Department of Telecommunication, Information Studies, and Media and Cognitive Science Program
Michigan State University
ASME DHM Conference 2007, Seattle, WA
Digital Ergonomics Meets Evolution

A need to account for phenotypic differences between individuals.

Key concept: homology.

Homology = similar structures present across individuals, but exist in different states.

* tree-like computational structures represent homologies well.

* need more information (genetic) to order these trees into a set of evolutionary relationships.
Digital Ergonomics Meets Evolution (con't)

Compare interspecific cases with intraspecific cases: divergence + functional polymorphism:

* look at divergence alone, human trees unresolved, results mimic demography.

* look at single nucleotide polymorphisms (SNPs) alone, no idea of how evolution acts upon gene networks or gene action.
Genotype-to-Phenotype Mapping in Nature

Environment

Phenotype

Not well understood

Regulatory/Physical Interactions

Heresy in some circles

!?!

Genotype
Constructing a Phenotype: using Genetic Algorithms (GAs)

Goal of genetic encodings: a phenotype that represents a range of variation (anthropometric end states):

* traits are polygenic. Large number of alleles per gene also produces non-isomorphic mappings between genotype and phenotype.

* regulatory effects; creates “continuous” traits (e.g. incremental increase or decrease in the length of a snout across fishes).

* what kind of “gene action” is required to produce traits that correspond to the 5\textsuperscript{th} percentile? 50\textsuperscript{th}? 95\textsuperscript{th}?
Constructing a Phenotype: using Genetic Algorithms (GAs) (con't)

Problem of Representation:

* Chromosome (all “loci” or “genes” involved in problem space).

* Gene (or locus): a series of binary values that encode some solution to a problem.

* Operators: mutation, recombination (macromutation), drift.

* Epistasis: interaction between genes as they are transformed into a phenotype (G-to-P mapping).
Constructing a Phenotype: using Genetic Algorithms (GAs) (con't)

Feedforward model of G-to-P mapping (representational output):

* assumes that epistasis is probabilistic, combinatoric.

* with or without regulation (simple vs. complex model).

* produces a series of phenotypes (population-level modeling) which can be classified phylogenetically.

* produces a range of variation that follows a distribution (nth percentile of phenotypic distribution can be assessed).
Constructing a Phenotype: Simple Epistatic Model
Constructing a Phenotype: Complex Epistatic Model
Encoding of Problem (statistical vs. in silico “QTL”)

This essentially becomes an inverse problem:

* database-driven representations (CAESAR).

* P-to-G encoding (observed cases). Parameterize problem.

* G-to-P mapping (expected/evolved distribution).

* encoding step similar to QTL (quantitative trait loci) analysis.
Encoding of Problem (statistical vs. in silico “QTL”) (con't)

Quantitative Trait Loci (QTL): method used for making the link between genotype and phenotype.

* breed a generation of controlled crosses.

* survey population for presence/absence of certain alleles (or sequences that resemble alleles).

http://www.nervenet.org/papers_images/cbQTLs.jpg
Encoding of Problem (statistical vs. in silico “QTL”) (con't)

In Japanese Black Cattle (Malau-Aduli et al, Animal Science J., 2005) and dogs (Carrier et al, Genome Research, 2005), QTL used to find the genetic components of upper- and lower-body segments:

* biologically, correlate phenotypic traits with presence of certain sequences (statistical associations).

* assume that sequences on a particular portion of the chromosome is the allele in question.

* in nature, controlled crosses ~ isolation of mechanism; in digital context, assumption replaced by making educated guesses about genomic state of phenotype.
Statistical vs. Biological Epistasis

Gene-Gene interactions (GxG):

* “polygenic” effects, pleiotropic effects (effects of many genes simultaneously and one gene/many products, respectively).

Gene-Environment interactions (GxE):

* endocrine (hormonal) triggers, physiological stressors, bone remodeling, facultative adaptation (e.g. perceptual accomodation).

Biological epistasis: GxG only, statistical epistasis: GxG + GxE.
Statistical vs. Biological Epistasis (con't)

The connections between units in the simple and complex model (matrix $w_{ij}$) represent the epistatic interactions.

* statistically, gene-gene interactions are conceptualized in this way for a particular trait.

* regions of the matrix yield additive and multiplicative epistasis.

* certain sets of interactions, “stable states” of phenotype (representing phenotypes from 5th to 95th percentile?)
Representation of “Genes” and “Alleles” \textit{in silico}

Sets of genes: \( \text{Gene A} = \{0,0,1,1,0\} \), \( \text{Gene B} = \{1,1,0,1,0\} \)

Two different allelic states for same gene = \( \text{Gene A} = \{0,0,1,1,0\} \), \( \text{Gene A}' = \{0,0,1,0,0\} \)

* Gene A and Gene B map to different combinations of phenotypic elements: (\( \text{Gene A} \gg P_1, P_3, P_4 \); \( \text{Gene B} \gg P_1,P_2 \)).

* Intergenic and intragenic variation; phylogenetic relationships can be discovered for single genes, phenotypic variance assessed across allelic states for a specific gene.
5th percentile (above) and 95th percentile (below):

* relatively small # of specimens (assuming trait size is randomly distributed).

* more if large-scale recombination takes place.

* “tails” of distribution show more variation (resequencing)?
Representation of “Genes” and “Alleles” *in silico* (con't)

50th percentile:

* large # of specimens if genes retain physical linkage in evolution.

* representative of computational “wild type” (mutant phenotype = representation of low frequency alleles in genotype).
Importance of Incorporating Regulatory Information in Model

Digital modeling of macromutation: large-scale change in genotype/phenotype system.

Harris et al. (Science, 316, 235, 2007) have developed a method for finding differences between taxonomic groups called “genomic triangulation”; can be used to model macromutation:

* align sequences (chromosomes) in a pairwise manner

* discover conserved regions (shared between taxa). Alignment “gaps” indicate “breakpoints”

* infer specific chromosomal rearrangements.
Importance of Incorporating Regulatory Information in Model (con't)

Importance of gene regulation in physiology and human performance:

* point mutations: do they code for different amino acids (Ka) or the same amino acid (Ks)?

* high Ka/Ks ratio: positive selection. Positive selection, evolutionary “change” in organ.

Figure 1. The Ka/Ks Ratios between Human and Mouse for Genes Expressed in Brain, Liver, and/or Muscle
The number of genes is given in parentheses above each Ka/Ks value; the 95% confidence interval is given in parentheses below each Ka/Ks value. Expression data for brain and liver were from Enard et al. [5], and data for muscle were from Public Expression Profiling Resource (http://pepr.cnmcresearch.org). doi:10.1371/journal.pbio.0050013.g001
Observations of Selection on Genotype

Bakewell et al (Nature, 2007): more positively selected genes (PSGs) found in chimps than in humans (based on protein sequences).

* 154 genes underwent positive selection in human lineage.

* 233 genes underwent positive selection in chimp lineage.

* unshared genes were involved in metabolism, transcriptional regulation, and stress response.

* relevance to epigenetic fitness function in model? Connections ($w_{ij}$) can be strengthened, weakened in different lineages to produce different results.
Role of “Divergence” vs “Polymorphism” data in model

In the previously reviewed result, functional “polymorphism” was more important than “divergence” data.

* in model: allelic states are “divergence”, while epistatic connections are “polymorphic” (gene action vs. regulation).

* modifying connection weights affect regulatory factors (fitness of various GxG and GxE interactions).

* modifying allelic states: how “different” will phenotype X be if state of gene X is changed from allele A to allele B? Distinct from regulatory factors, even in polygenic mappings.
Beyond the 95\textsuperscript{th} percentile: simulating the results of “resequencing” studies.

A recent study (Ahituv et al, American J. Human Genetics, 2007) examined heritable variation related to the body-mass index (BMI - 5\textsuperscript{th} percentile and below 95\textsuperscript{th} percentile and above):

* most genetic sequences from population represent “average” phenotypes.

* solution: sequence candidate genes from phenotypes on the tails of the distribution (n = 379 "obese" and 378 "lean").

* greater number of alleles /polymorphisms are expected for this portion of the population.

* also maps to a greater variance in phenotype.
Toy Problem: “genetic” mapping of human hand

1) use as few genetic, phenotypic elements as possible.

* one gene, regulatory element produces self-similar elements, the other can be tuned to approximate naturally-occurring variation.

2) use as many genetic elements as possible (isomorphy between genotype-phenotype).

Two approaches (a continuum) for encoding basic human hand morphology:

<table>
<thead>
<tr>
<th>P-level feature</th>
<th>G-level representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Digit, proximal</td>
<td>Gene j x Gene k x Environment l</td>
</tr>
<tr>
<td>First Digit, distal</td>
<td>Gene i x Gene k x Environment i</td>
</tr>
<tr>
<td>Second Digit, proximal</td>
<td>Gene j x Gene k</td>
</tr>
<tr>
<td>Second Digit, distal</td>
<td>Gene i x Environment k</td>
</tr>
<tr>
<td>Third Digit, proximal</td>
<td>Gene j x Gene k</td>
</tr>
<tr>
<td>Third Digit, distal</td>
<td>Gene i x Environment j</td>
</tr>
<tr>
<td>Fourth Digit, proximal</td>
<td>Gene j x Gene k</td>
</tr>
<tr>
<td>Fourth Digit, distal</td>
<td>Gene i x Environment h</td>
</tr>
<tr>
<td>Fifth Digit, proximal</td>
<td>Gene j x Gene k</td>
</tr>
<tr>
<td>Fifth Digit, distal</td>
<td>Gene i x Environment i</td>
</tr>
</tbody>
</table>
Toy Problem: “genetic” mapping of human hand (con't)

24-bit color palate using the infinite alleles model (256 alleles, 3 genes, and allow for random mutation):


**Genotype A:**
```
1,1,0,0,0,1,0,1,1,0,0,1,1,1,0,1,0,0,1,0
```

**Genotype B:**
```
1,1,0,0,0,1,0,1,0,0,0,1,1,1,0,0,1,1,1,0
```

**Genotype (cross):**
```
1,1,0,0,0,0,1,0,1,0,0,0,1,0,1,0,1,1,1,0,0,1,0
```

![Hand illustrations](Image1.png)
Goal: towards a “human analogue”

Why bother? One goal is to create a human analogue.

* test theories of human evolution *in silico* (can’t do functional assays *in vivo* or observe evolutionary processes directly, but can build digital models informed by biology).

* test the potential range of human variation for a trait. We can’t measure everyone, but we can extrapolate empirical distribution of phenotype (discovery of “latent” variation).

* perform population-based DHM. Using computational methods, we can “evolve” alternate phenotypes. This can be used to approximate ethnic differences or disease states.
Potential Application Domains

Telepresence:

* Magenaat-Thalmann et al (J. Computer Science and Technology, 2004) have parameterized and classified major features of a virtual mannequin using CAESAR.

* these data are currently used for purposes of “trying on” clothing purchased through an e-commerce website.

Biometrics:

* discover the regularities inherent in fingerprints, retinal surface, and facial features.

* deal with variation found in both discrete groups and outliers (sample of normal and abnormal phenotypes).

* model results could be used as data to train a system against generated attacks.
Potential Application Domains (con't)

Ergonomics:
* GENESIS models (Seo et al, IEEE Computer Society, 2002) are populations of avatars augmented by key parameters.
* epistatic models may be a way to customize members of these populations to represent human variation found in nature.

Forensics and Drug Discovery:
* facial/skeletal reconstruction. Landmarks and soft tissue depth could be estimated (Domaracki and Stephan, J. Forensic Science, 2006).
* surgical planning: predicting what phenotypic traits look like after modification/surgery.
* estimating disease phenotypes or the evolution of the human immune system.