Vertebrate Limb Chondrogenesis and Invertebrate Comparative Genomics

Presented by

Scott Christley
Department of Computer Science and Engineering
Department of Mathematics
Interdisciplinary Center for the Study of Biocomplexity
Center for Global Health and Infectious Diseases
University of Notre Dame

Wayne State University, April 7th 2008
Presentation Outline

- Vertebrate Limb Chondrogenesis
  - Biological Background
  - Computational Model
  - Results
  - Future Directions

- Population Ecology and Climate Change
- Invertebrate Comparative Genomics
  - VectorBase.org
  - Ultraconserved Elements
  - Anopheles gambiae PEST, M and S
- Free/Open Source Software Community (if time permits!)
  - Data Mining, Social Network Analysis
Vertebrate Limb Chondrogenesis

- **Morphogenesis**: Formation of the structure of an organism or part; differentiation and growth of tissues and organs during development.
- **Organogenesis**: The formation and development of the organs of living things.
- **Limb Chondrogenesis**: The differentiation of mesenchymal cells in the developing (vertebrate) limb into cartilage as the primordia of skeleton.

- NSF (IBN-0344647 and EF-0526854)
- Notre Dame Biocomplexity Cluster (NSF MRI Grant No. DBI-0420980)
- Center for Applied Mathematics
Skeletal Pattern Formation: stages in a chicken limb bud
In vitro cell-culture experiments

Leg condensations visualized by Hoffman Contrast Modulation optics after 48 hours.

Leg condensations visualized by Hoffman Contrast Modulation optics after 72 hours.
Additional Biology

• Exposure of cells to TGF-β causes immediate production of fibronectin mRNA.

• FGFs produced by limb ectoderm induce a perinodular (i.e., lateral) inhibition of condensation formation that depends on the transient presence of FGF receptor 2 (FGFR2) at sites of condensation.
Hypothesized Model

- Reaction-diffusion (RD) system with a TGF-β-type morphogen as activator and an FGF-induced lateral inhibitor.
- RD system forms spatial patterns of activator peaks and inhibitor valleys.
  - Induces fibronectin (FN) mRNA production at activator peaks and inhibits at inhibitor valleys.
Computational Model

- Discrete, multiscale stochastic agent-based model
- Agent-based cell representation
  - Extended, multi-pixel cells that maintain internal state throughout their spatial extent.
  - Follow stochastic simple rules to change shape, move, respond to molecule (differentiation), and interact with fibronectin.
- Separation of physical scales for cells (coarse grid) and molecules (fine grid).
- Multiple temporal scales (diffusion, chemical reactions).
- Implement discrete stochastic reaction-diffusion system.
- Implemented in Objective-C (object-oriented C language variant).
Cellular Model

- Cells have essentially isotropic geometry, that is they do not elongate in the direction of migration but rather probe their environment by extending short randomly placed projections.
- The cell nucleus is also isotropic but is relatively unchanging in shape and comprises more than half the cell volume.
Cellular Model

- Cells in fibronectin-rich, condensing areas of the micromass round up such that their cross-section in the plane of the culture is reduced by 20-30 percent.
Reaction-Diffusion


- Discrete stochastic implementation.
- Reaction: activator (U), inhibitor (V)
  - \[ \Delta U_t = \min\{ \text{MAX}_U, (k_1 + B_U)U_t \phi_t + k_2 V_t \} \]
  - \[ U_{t+1} = \max\{ 0, U_t + \text{round}(\Delta U_t) \} \]
  - \[ \Delta V_t = \min\{ \text{MAX}_V, k_3 U_t \phi_t + k_4 V_t \} \]
  - \[ V_{t+1} = \max\{ 0, V_t + \text{round}(\Delta V_t) \} \]
  - \( k_1 \) and \( k_3 \) positive, \( k_2 \) and \( k_4 \) negative

- Diffusion
  - random walk (up, down, left, right)
  - \[ D = p \cdot n \quad 0 < p < 1 \]
  - \( D_V > D_U \)

```
Algorithm 3 calculateReactionDiffusion()

Calculate chemical reaction for each pixel on grid.
for i = 1 to n do
    Calculate activator and inhibitor diffusion for each pixel on grid.
end for
```
Main Simulation

Algorithm 4 Main Simulation

\begin{algorithm}
\begin{algorithmic}
\For{each simulation iteration do}
\State Generate randomized list, \( R \), of agents.
\For{each agent in \( R \) do}
\State moveWithProbability\((p)\)
\State changeShape()
\EndFor
\State calculateReactionDiffusion()
\State Determine if any cells have reached threshold for differentiation.
\State Calculate fibronectin production for each differentiated cell.
\EndFor
\end{algorithmic}
\end{algorithm}
## Model Calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Physical Value</th>
<th>Simulation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell diameter/area</td>
<td>15um / 177um²</td>
<td>7 pixels</td>
</tr>
<tr>
<td>Cell spatial grid</td>
<td>1.4 x 1.0 mm</td>
<td>280 x 200 pixels</td>
</tr>
<tr>
<td>Molecular spatial grid</td>
<td></td>
<td>560 x 400 pixels</td>
</tr>
<tr>
<td>Spatial ratio cells/molecules</td>
<td>10000 : 1</td>
<td>28 pixels : 1 pixel</td>
</tr>
<tr>
<td>Simulation Temporal scale</td>
<td>17.07 sec</td>
<td>1 iteration</td>
</tr>
<tr>
<td>Reaction Temporal scale</td>
<td>17.07 sec</td>
<td>1 reaction</td>
</tr>
<tr>
<td>Diffusion Temporal scale (n = 200)</td>
<td>85.3 msec</td>
<td>1 diffusion step</td>
</tr>
<tr>
<td>Basal activator production (B_u)</td>
<td>unknown</td>
<td>28</td>
</tr>
<tr>
<td>Activator self-regulation (k_1)</td>
<td>unknown</td>
<td>0.3356</td>
</tr>
<tr>
<td>Activator regulation of inhibitor (k_3)</td>
<td>unknown</td>
<td>0.16</td>
</tr>
<tr>
<td>Inhibitor regulation of activator (k_2)</td>
<td>unknown</td>
<td>-1.1</td>
</tr>
<tr>
<td>Inhibitor decay (k_4)</td>
<td>unknown</td>
<td>-0.4615</td>
</tr>
<tr>
<td>Maximum activator produced (MAX_u)</td>
<td>unknown</td>
<td>8000</td>
</tr>
<tr>
<td>Maximum inhibitor produced (MAX_v)</td>
<td>unknown</td>
<td>8000</td>
</tr>
<tr>
<td>Cell differentiation threshold (CDT)</td>
<td>unknown</td>
<td>7000</td>
</tr>
<tr>
<td>Activator diffusion rate (D_u)</td>
<td>10 um² /sec (1)</td>
<td>27 pixels/iteration</td>
</tr>
<tr>
<td>Inhibitor diffusion rate (D_v)</td>
<td>unknown</td>
<td>108 pixels/iteration</td>
</tr>
<tr>
<td>Cell diffusion rate</td>
<td>0.42 um² /min</td>
<td>1 pixel/60 iterations</td>
</tr>
<tr>
<td>Cell diffusion rate on fibronectin</td>
<td>0.62 um² /min</td>
<td>1 pixel/40 iterations</td>
</tr>
</tbody>
</table>

Results

Reproduce Condensation Patterns

(A) Discrete spot-like precartilage condensations in a micromass culture of limb mesenchymal cells. (B) Spatial grid of equal physical size to (A) containing over 6000 cells produced by simulation showing clusters of fibronectin-producing differentiated cells (white), non-differentiated cells (blue gray), and empty space between cells (black). (C) Spatial grid of fibronectin-rich patches (black). (D) Activator concentration.
Variation in some of the key parameters induces morphological changes in the resultant spatial patterns from distinct spots to connected spots to stripe-like patterns. Average peak interval versus average island size for variations in some of the key parameters are shown: +5% (diamond) and -5% (filled diamond) for activator self-regulation ($k_1$), +5% (triangle) and -5% (filled triangle) for activator regulation of inhibitor ($k_3$), +5% (down triangle) and -5% (filled down triangle) for inhibitor regulation of activator ($k_2$), +5% (plus) for inhibitor decay ($k_4$). The colored points are a gradient of variations: 1% (red), 2% (orange), 3% (green), 4% (blue), 5% (violet). Also shown are the five simulations (square) using the standard parameter values and the mean for the twelve experiments (circle).
(A) Stripe-like precartilage condensations. (B) Spatial grid containing over 6000 cells produced by simulation showing stripes of fibronectin-producing differentiated cells (white), non-differentiated cells (blue gray), and empty space between cells (black). (C) Fibronectin-rich stripes (black) produced by the differentiated cells. (D) Activator concentration.
Alternate Hypotheses

- Two dynamical regimes can produce condensation patterns
  - Oscillatory
  - Stationary

- An important implication is that developmental processes do not require a strict progression from one stable dynamic regime to another, but can occur by a succession of transient dynamic regimes tuned (e.g., by natural selection) to achieve a particular morphological outcome.
Future Directions

• **Question of the diffusible lateral inhibitor**
  – Has not been experimentally found

• **Juxtacrine signaling (Notch-Delta pathway) as alternative mechanism for long-range inhibitor.**
Juxtacrine Signaling

Diagram showing the signaling process involving genes such as hes1, Delta, and Notch.
Summary

• Biology
  – Calibrate model with known experimental values.
  – Reproduce experimental data.
  – Sensitivity analysis of key parameters reveals morphological variation.
  – Show that spot and stripe patterns are close in parameter space.
  – Disclose two distinct dynamics regimes, transient and stationary, that suggests biological hypotheses that can be empirically tested.

• Selected publications


Population Ecology and Climate Change

- **Phosphoglucone Isomerase (PGI)**
  - Key enzyme in glycosis --> Flight
  - Balancing selection
    - Heterozygote is most fit
    - One genotype: kinetic effectiveness
    - Other genotype: thermal stability
- How will climate change affect genotype frequencies?
- Can species adapt fast enough?

Jessica Hellmann  
University of Notre Dame

Ward Watt  
Stanford University

Mark Alber  
University of Notre Dame

Richard Gejji  
University of Notre Dame

Colias meadii male
Presentation Outline

• Vertebrate Limb Chondrogenesis
  – Biological Background
  – Computational Model
  – Results
  – Future Directions
• Population Ecology and Climate Change

• Invertebrate Comparative Genomics
  – VectorBase.org
  – Ultraconserved Elements
  – Anopheles gambiae PEST, M and S

• Free/Open Source Software Community
  – Data Mining, Social Network Analysis
VectorBase.org

NIH/NIAID Bioinformatics Resource Center for Invertebrate Vectors of Human Pathogens

Frank Collins
University of Notre Dame

Greg Madey
University of Notre Dame

*Culex pipiens* (West Nile)

*Anopheles gambiae* (Malaria)

*Ixodes scapularis* (Lyme disease)

*Pediculus humanus* (Trench fever)
VectorBase.org Architecture
Ultraconserved Elements


- 100% identity (no insertions, deletions) of nucleotides between two or more genomes.
- 481 segments longer than 200 bp between human, mouse, and rat. (exclusive rRNA)
  - Only 1/4 overlap known mRNA
  - 1/2 no known transcription
  - Remaining 1/4 has inconclusive evidence.

- Why such extraordinary conservation?
- 1.2% of the human genome codes for proteins, what is the rest for?

- Elements extracted from whole genome alignment.
  - Alignments computationally expensive
  - Difficult for evolutionarily distant organisms

Algorithm and Workflow

Suffix Array

- **Suffix Trees**
  - MUMmer
  - Large memory requirement

- **Suffix Arrays**
  - Smaller
  - Disk storage
  - Algorithms are just as efficient with enhancements

Multiple Organisms

17 Vertebrates
- Human
- Mouse
- Rat
- Dog
- Cow
- Chicken
- Chimpanzee
- Macaque
- Elephant
- Rabbit
- Armadillo
- Tenrec (hedgehog)
- Opossum
- Tetraodon (pufferfish)
- Zebrafish
- Fugu (pufferfish)
- X.tropicalis (frog)

About one week computation time on Biocomplexity cluster (8 computers).

Anopheles gambiae PEST, M & S

Nora Besansky
University of Notre Dame

Scott Emrich
University of Notre Dame

• Anopheles complex is undergoing recent speciation
  – Incipient species, M & S molecular forms
  – Current genome (PEST) is mixture of M & S

• Comparative genomics
  – Because genomes are so similar, characterize differences instead of conservation
  – Hypothesis generation of events indicative of speciation

• Transposable element discovery and annotation pipeline
  – Annotate individual elements in M & S genomes
Open Source Software (OSS) - Linux

- Free …
  - to view source
  - to modify
  - to share
  - of cost

- Examples
  - Apache
  - Perl
  - GNU
  - Linux
  - Sendmail
  - Python
  - KDE
  - GNOME
  - Mozilla
  - Thousands more
OSS Activity Network

- Dataset: SourceForge.net database
  - 168,000+ registered projects 1,786,000+ registered users

- User/Developer performs an activity for a project.
- 29 activities; submit bug, submit feature request, assign bug, post forum message, create file release, add/modify source code, etc.
  - 120+ million activity events
- Multi-relational, weighted, bipartite network.
  - Activity = relation, weight = activity count
- Algorithm for clustering with non-parametric statistical test
  - Activity distribution for user/project pair defines a sample for our statistical test.
Social Positions

Administrator

Developer

Handyperson

Message Poster
Acknowledgements

• Dr. Stuart Newman, New York Medical College
• Dr. Greg Madey (Computer Science co-advisor)
• Dr. Mark Alber (Mathematics co-advisor)
• Dr. Frank Collins, Dr. Nora Besansky, Dr. Scott Emrich

Funding

• NSF (IBN-0344647, EF-0526854 for limb), (0222829 for OSS)
• Notre Dame Biocomplexity Cluster (NSF MRI Grant No. DBI-0420980)
• Notre Dame Center for Applied Mathematics
• Notre Dame Center for Global Health & Infectious Disease, and the NIH/NIAID under Contract No. HHSN266200400039C for “VectorBase: A Bioinformatics Resource Center for Invertebrate Vectors of Human Pathogens” [www.vectorbase.org]
Thank You!