Modelling Leaf Shape Evolution

Jotun Hein, Nick Jones, Miltos Tsiantis, James Anderson and John Moriarty. 17.6.2011

Background and Motivation. The foundations of evolution and population genetics respectively can be found in studies at the level of the organism. Such models will virtually always focus on intrinsically discrete features/characters such as nucleotides or the absence-presence of a character capturing a continuous feature. However, this precludes a large class of characters that are of major interest in biology and cannot be well reduced to discrete characters, namely shapes and patterns. This project will only focus on simple models of shape evolution and their application to leaf shapes.

Shape analysis is a well established, but difficult, topic that is approached through a variety of models. Shape/form/patterns are considerably more complicated than other biological because it cannot be constructed by a cartesian product [vector] of real numbers, integers, set or group and thus the standard concepts describing relationship between entries in a vector, such as correlation or dependence are insufficient as shape can be a highly structured object.

The specification of a shape might involve both discrete components [for instance a topological classification] and continuous parameters like sizes, lengths etc. The latter implies that there often is a transfinite number of shapes and it is rare to observe a transfinite number of variables. Often a shape is then estimated by observing/measuring landmarks on the shapes. This is sometimes seen as a necessity but can also be a source of error, since choosing these landmarks are not perfect. New methods attempt to avoid the specification of landmarks and, in principle, operate on functional data (Jones and Moriarty, 2011 Ramsay and Silverman 1998) .

Shape analysis can address a variety of important questions: 1) Inference of shape from incomplete data [e.g. part of a skull can be missing]: 2) Inference of an evolutionary tree given (homologous) data from different nodes of that tree 3) Inference about ancestral shapes given both evolutionary tree and samples at its nodes.

The real situation will often complicate the situation even further, by having additional non-shape data that is important in explaining the shape. The largest class of extra information is, of course, genomic.

There are some models of shape/gradients that explains the shape/gradient in terms of explanatory variables such as promoters/gene regulation (for instance Janssen et al, 2006) or networks (for instance Mjollsness, 1991). These are not widely accepted models, but rather designed to explain a given shape with focussed data obtained from a hypothesis.

The Data and Biological Problem – Leaves of seed plants show tremendous variation in shape, making them an attractive system to study the evolution of morphological diversity. Such study is a main focus of research in the Tsiantis group (Piazza et al, 2010; Hay and Tsiantis, 2006). Leaf shape can be described as simple, if the blade is undivided, or compound, if the blade is divided into distinct leaflets, which are sometimes themselves sub-divided to additional orders of leaflets. Further, leaflets and leaves can be smooth or have marginal outgrowths of different depths and arrangements, such as lobes or serrations. This vast variation in leaf form offers exciting opportunities to explore the genetic basis for diversification of organ geometry.

Here we will study leaf shape in six crucifer plant species (Arabidopsis arenosa, Arabidopsis halleri, Arabidopsis lyrata, Olimarabidopsis pumila, Arabidopsis suecica, Arabidopsis thaliana) that shared a common ancestor approximately 20 million years ago. Current evidence indicates that within this evolutionary lineage there was gradual simplification in leaf morphology. During this simplification process a fully dissected leaf with up to 4 pairs of leaflets gave rise to a deeply lobed simple leaf, which then underwent a further shape transition to an unlobed leaf form exemplified by the

model organism *Arabidopsis thaliana*. The leaf shapes of the six species we will study represent all three types of leaf morphologies discussed i.e. fully dissected, lobed, and unlobed. We will develop and compare methodologies aiming to accurately capture and quantify leaf diversity in a representative sample of our species. This information together with knowledge about the phylogenetic relationships of the species we study will allow us to mathematically approximate and understand the precise shape transitions that produced divergent leaf shapes during crucifer evolution.

A. suecica   A. thaliana   A. halleri   A. arenosa   O. pumila   A. lyrata   C. hirsuta
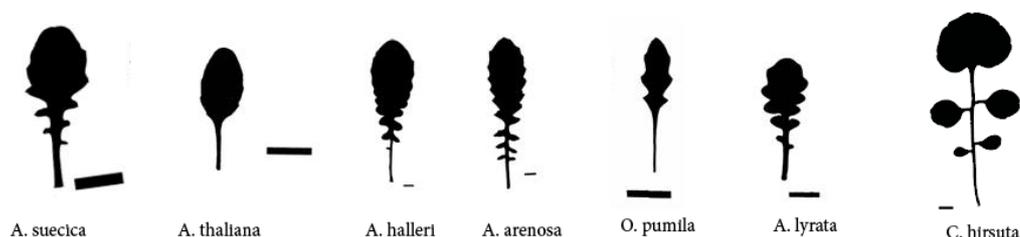
The data consists of several samples of each leaf outline (as above) with a known evolutionary tree. The leaf shape outlines will need to be converted to a set of data points from which the 1d Gaussian Process, used to model the evolution, can be inferred. This is well established as a procedure (for example, Rasmussen and Williams (2006)).

A key challenge to be solved is that leaf shapes can vary considerably owing to the presence of different numbers or orders of repeated shape elements such as lobes or leaflets. Such differences make it difficult to utilize conventional shape analysis methods many of which rely on investigating covariance of homologous parts within a structure. We will therefore aim to quantify leaf shape variation without the strict need to fix homologous points. Rather we will base our study on investigating the global properties of the curves that correspond to our leaf outlines.

Models to be used

The starting point will be the Jones and Moriarty (2011) model. This might be called "evolutionary functional data analysis" and extends sequence phylogenetics to functions (with co-domains of arbitrary dimension) by explaining how Gaussain Processes (like Random Walks and the Ornstein-Uhlenbeck process) can be put on phylogenies (it can also be viewed as a non-parametric approach to shape evolution). We will make a simple approximation that (after eliminating symmetries) the leaf edge can be treated as a function on 1-d which has been wrapped around a canonical leaf shape.

While treating leaf shapes as simple functions is a start, it clearly fails to capture the fact that biology can allow repeat units (developmental patterning though KNOX-genes – relatives of hox genes). To deal with this, we suggest a fusion of two approaches: a land-mark-free functional approach (to deal with smaller fluctuations in leaf shape) and one involving discrete characters (to capture the repeat patterning allowed by developmental genes). The students will explore both drastic changes in gene activity such as presence or absence of gene expression in leaves and subtler gene expression modifications within the functional framework. Such an approach does not have precedent and would be very cutting edge (it would also be useful for the study of protein shape evolution).

An alternative route is to really treat the leaf as a shape (rather than as a function) and help develop new land-mark-free and Bayesian non-parametric techniques for shape evolution. This is also a brand-new area and involves fusing the approaches in Holm et al. (2005,2009) with those in Jones and Moriarty (2011).

The project

The project would explore the applications of the above models to leaf shapes, by considering the following questions.

1. ***Can the model reconstruct the known evolutionary tree from the leaf shape data?*** Jones and Moriarty (2011) develop methods for constructing evolutionary trees given an underlying Gaussian Process evolutionary model. Clearly the model will be inappropriate for the problem if the evolutionary tree reconstructed is a poor representation of the true evolutionary tree.

2. ***What do the inferred ancestral leaves look like?*** Do these lie within biological reason? How do these leaf shapes change when using the true evolutionary tree as opposed to the inferred evolutionary tree? Again, this would also give insights to the appropriateness of the model.
3. ***How stable is the model?*** The data contains several samples of each leaf type. How do the inferred evolutionary tree and corresponding ancestral leaves change with the different samples? In the same vein, how much noise can be permitted in the data to still get strong inferences?

Looking at these questions will hopefully provide insights into both the process of leaf shape evolution and the model of Gaussian Processes for shape evolution modelling.

Project Plan

-Organise the data (leaf shapes and phylogenies) and undertake appropriate reading.
-Implement a simple version of the Jones and Moriarty (2011) algorithm and apply to leaf edges.-
Consider extensions:

      1) possible integrations of evolutionary functional data analysis with discrete characters.
      2) consider leaf shape evolution as a trajectory in shape space and make connections with the work of Holm et al. (2005,2009).

Skills needed – the project uses biology (ability to read literature about molecular biology of leaves), statistical and stochastic modelling and ability to efficient implementation of models/simulations.

References

Key references
Gaussian Processes
[1] Jones and Moriarty, (2011). Evolutionary Inference for Functional Data: Using Gaussian Processes on Phylogenies to Study Shape Evolution. http://arxiv.org/abs/1004.4668
[2] Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning, Sections 2.1-2.3.* MIT Press. http://www.gaussianprocess.org/gpml/chapters/RW.pdf
[3] http://videolectures.net/gpip06_mackay_gpb/
[4] http://videolectures.net/mlss07_rasmussen_bigp/

Leaf Shape Evolution
[5] Piazza et al., (2010). Arabidopsis thaliana Leaf Form Evolved via Loss of KNOX Expression in Leaves in Association with a Selective Sweep. *Current Biology* **20**, 1-6.
[6] Hay and Tsiantis, (2006). The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta. *Nature Genetics* **38**, 942-7.
[7] Klingenberg (2010). Evolution and development of shape: integrating quantitative approaches. *Nat. Rev. Genet.* **11,** 623–635.
[8] http://www.plantcell.org/site/teachingtools/TTPB2.xhtml
[9] http://www.plantcell.org/site/teachingtools/TTPB3.xhtml

Supplementary references

Gunz (2009) Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario
Comparison of cranial ontogenetic trajectories among great apes and humans Philipp Mitteroeckera*,
Janssens et al. (2006) Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even skipped* gene *Nature Genetics* - 38, 1159 - 1165
Grenander and Miller (2007) Pattern Theory - From Representation to Inference OUP

Jonke et al. (2007) Secular trends in the facial skull from the 19th century to the present, analyzed with geometric morphometrics American Journal of Orthodontics and Dentofacial Orthopedics 132.1.63-70
Klingenberg (2010) " Evolution and development of shape: integrating quantitative approaches" Nat. Rev. Genet. 11, 623–635
Lockwood et al.(2004) Morphometrics and hominoid phylogeny: Support for a chimpanzee–human clade and differentiation among great ape subspecies PNAS 101.13 4356–4360
Mardia and Dryden (1998) Shape Analysis Wiley
Mjolsness E, Sharp DH & Reinitz J (1991). A connectionist model of development. J Theor Biol 152: 429-453
Murray (2002), *Mathematical Biology 4th ed.,* Springer

Evolutionary Morphing David F. Wiley

Polly, D 2008, "Developmental Dynamics and G-Matrices: Can Morphometric Spaces be Used to Model Phenotypic Evolution? "Evol. Biol. 35, 83-96

Turing, *The chemical basis of morphogenesis,* Phil. Trans. Roy. Soc. Lond. B237, 37 (1952)

Wolpert (1969) Positional information and the spatial pattern of cellular differentiation. J.Theoret.Biol., 25, 1 47325 36

http://www.isbe.man.ac.uk/~bim/

http://graphics.idav.ucdavis.edu/research/projects/EvoMorphD. Holm, T. Ratnanather, A. Trouvé, L. Younes: Soliton Dynamics in Computational Anatomy, *Neuroimage* (2005) D. Holm, A. Trouvé and L. Younes . "The Euler Poincare Theory of metamorphosis". Quartely J. Applied Mathematics (2009)