

Foundations of the Age-Area Hypothesis

Matt Baker

Background

- The economic basis for indigenous institutions:
 - Baker (2003, 2008), Baker and Miceli (2005), Baker and Jacobsen (2007, 2008).
- Exploring the relationship between environment, technology, and institutions.
- Cross-cultural data sets
- Interesting perhaps, but of limited larger interest...

But recently...

- Applications in economic growth:
 - Alesina et. al. (2005), Spolaore and Wacziarg (2013), Michalopoulos (2012), Fenske (2012)
- Computational linguistics and Phylogenetic approaches to analyzing cultural diversity (Mace, 2006)
 - A computational field blending tools from biology, geographical data.
- Incorporation of geographical data into analyses

Question: How did ethnic and geographic diversity that we observe today come about?

Difficulties

- Galton's problem - Dependencies between cultures
 - Problems of inheritance (vertical transmission)
 - Borrowing (horizontal transmission)
 - Behavioral ecology / adaptation

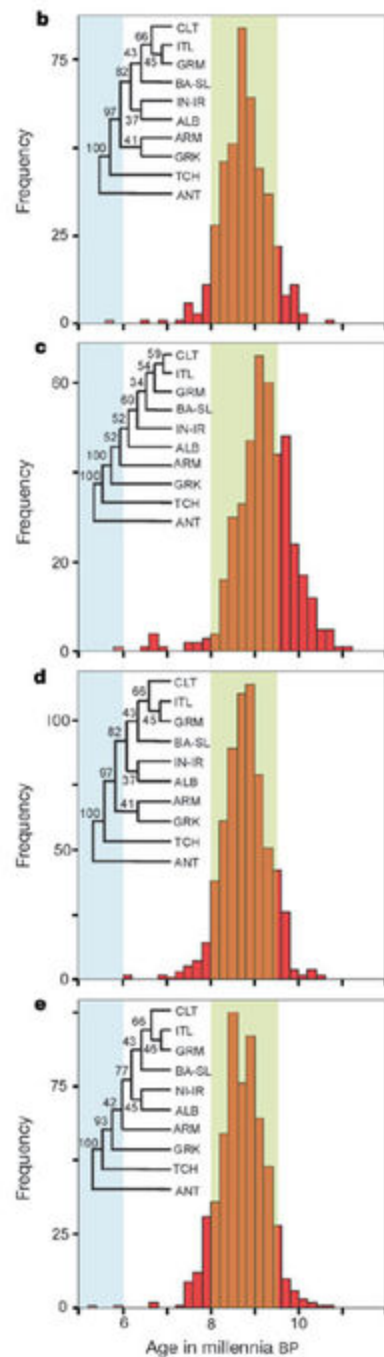
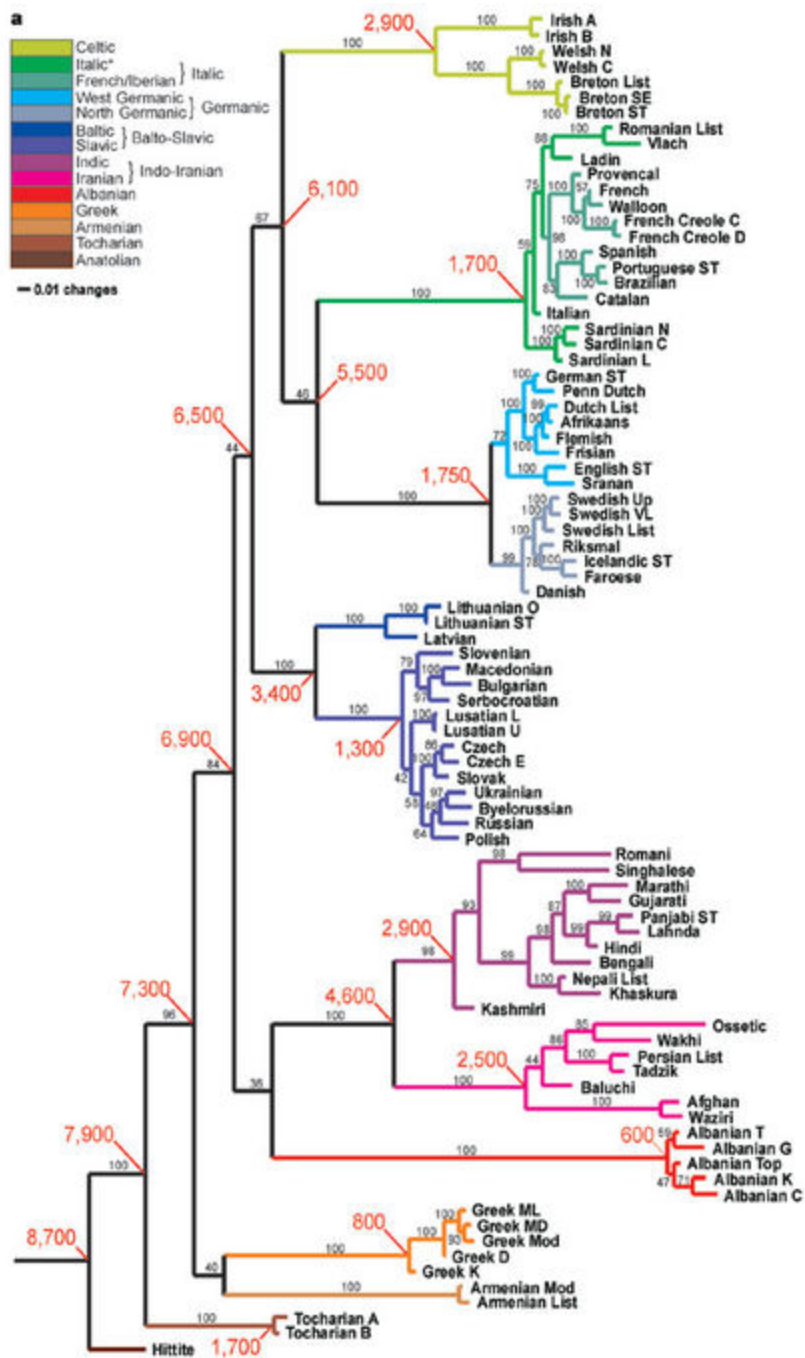
Difficulties

- Galton's problem - Dependencies between cultures
 - Problems of inheritance (vertical transmission)
 - Borrowing (horizontal transmission)
 - Behavioral ecology / adaptation

How can one know times and places?

Cultural Phylogenetics

- Modeling cultural evolution using Phylogenetic tools.
- Attractive and novel approach - explicit consideration of path dependency. Typical Econometric treatment a hammer in search of a nail?
- Computational linguistics - direct means of phylogeny-building (Mace, 2006)
- Atkinson and Gray (2006) example: Indo-European Tree.



Practical Questions:

- A sophisticated statistical description of Phylogenetic relationships between cultures.
- What can be said about the geography underlying the Phylogeny?
- How did the cultures on the tree come to be where they are?
- *Which related cultures have been in close proximity, and for how long? Who learned what from whom?*

The Age-Area Hypothesis (AAH)

- Sapir (1916) - the root of the Phylogenetic tree is the most likely geographical point of origin.
- Also: maximum divergence, maximum variety, maximum differentiation...
- Recursive application - migratory routes
- When coupled with a Phylogeny - we now have a details of time *and* place.
- Used to resolve historical debates, but also could be important in creating new theories

Old applications and continuing debates

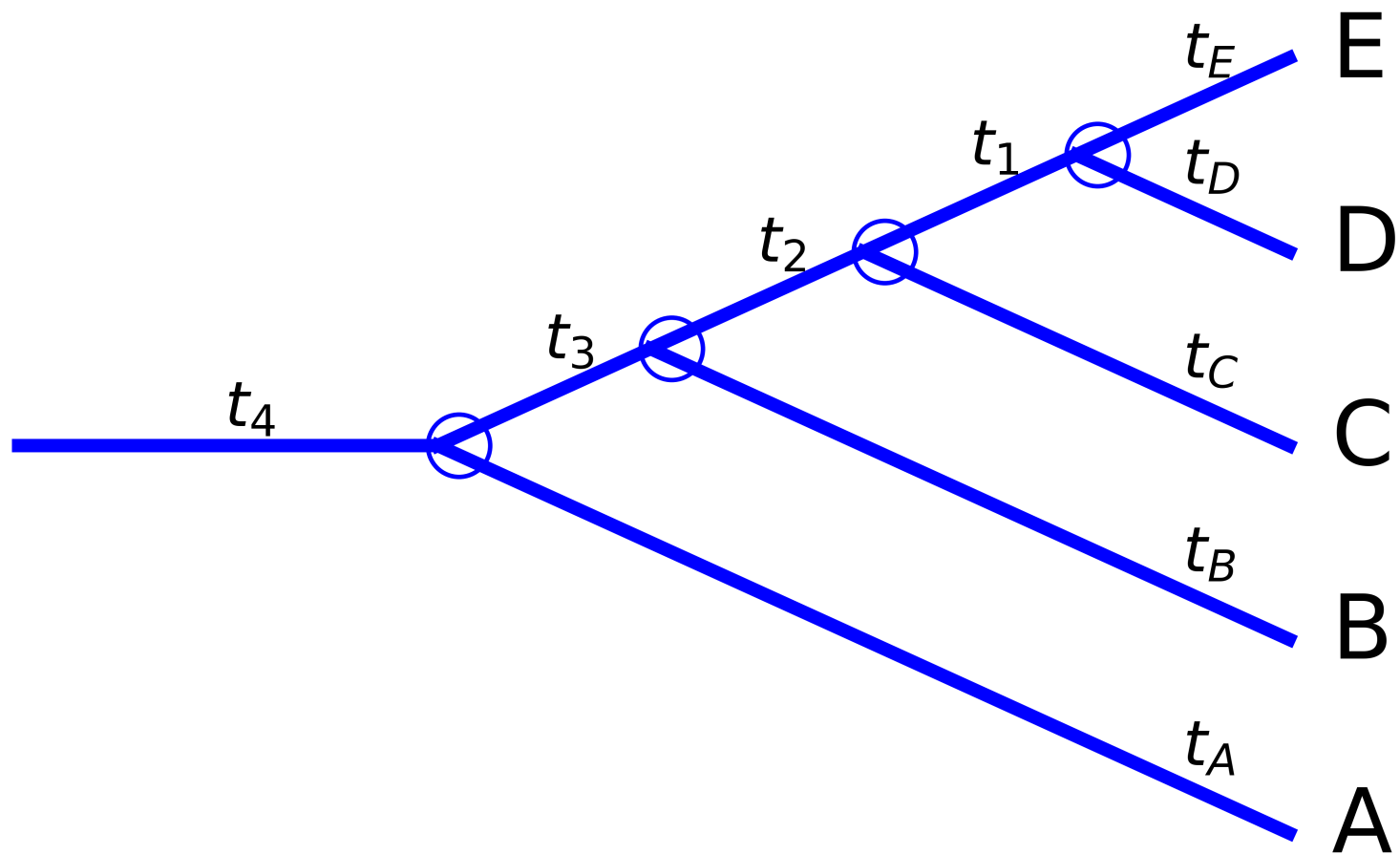
- Origins of Athabaskan/Na-Dene speakers
- Indo-European origins
- Afro-Asiatic origins
- Spread of Bantu peoples
- Native American population dispersal

On the need for theory...

- Greenhill and Gray (2005) write: "many expansion scenarios are little more than plausible narratives. A common feature of these narratives is the assertion that a particular line of evidence (archaeological, linguistic, or genetic) is 'consistent with' the scenario. 'Consistent with' covers a multitude of sins."
- Regressions and Spatial Econometrics - Leave me wondering what the DGP is...

So why believe the AAH (or not)?

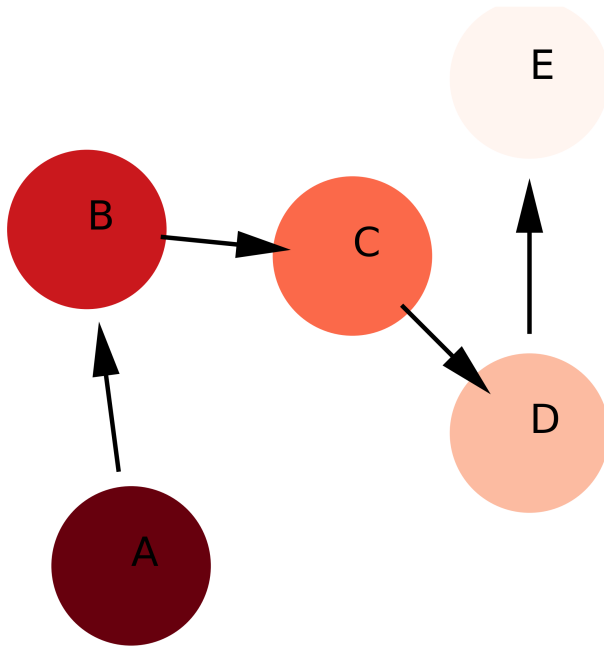
- Occam's Razor
- Minimum effort or # of moves
- Dyen (1956, p. 613) hits upon the idea of conserving moves of a particular sort: "...the probabilities of different reconstructed migrations are in inverse relation to the number of language movements required."



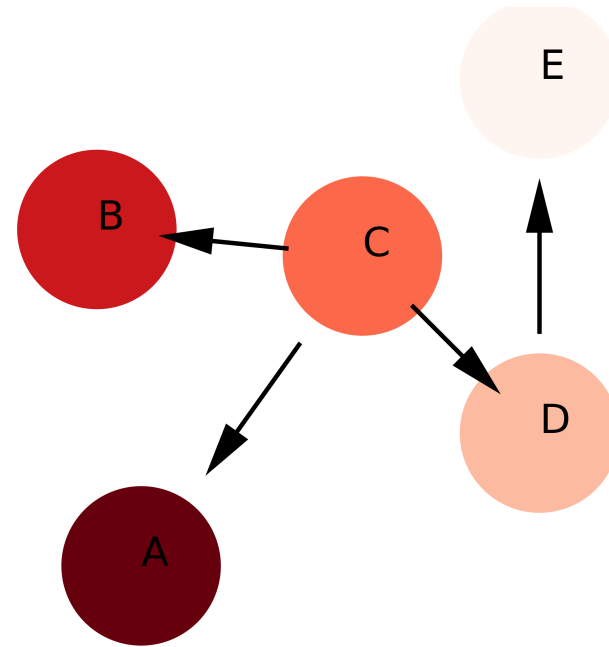
Problem Preview

Two Hypothetical Migratory Routes

A is point of origin



C is point of origin



Candidate Migratory Histories:

- A is point of origin - A to B to C to D to E
- C is point of origin - C to A, C to B, C to D to E
- Both are consistent with observed phylogenetic difference or drift. **The tree constrains the set of possible migrations**
- Note "minimum moves" doesn't get us very far. Both have four moves
- Actually - example approximates the debate between Ehret (2004) and Bellwood and Diamond (2003) about the origins Afrasan or Afroasiatic cultures/languages.

Basic Model:

- Assume a full, rooted binary tree
 - k terminal nodes/taxa/cultures, $k - 1$ internal nodes.
 $k - 1$ moves needed to span the tree.
- Current locations coincide with historic locations
- All constituents of the tree observed

Definitions

Migratory Event

A location jump from one location to a new, unoccupied one. For simplicity, the jump takes zero time and can be of any distance.

Migratory Chain

A chronological sequence of jumps through connected nodes that end at a terminal node/taxa/culture.

Migratory History

A collection of chains spanning the whole tree.

Basic assumptions

1. A migratory chain occupies one location at a time ("propensity to migrate" passed location to location).
2. A chain corresponds with a population movement. When a chain moves from its location to a new one, a new chain starts in its place.
3. Migratory chains move to new locations at random times, according to an Exponential/Poisson density.
4. Each migratory chain is unique in that it has its own parameters.

Chain One:

- requires a chain from A to B to C to D to E (or E to D)
- By the previous rules, new chains start at A, B, C, and D. Let T denote the length of the tree.
- Likelihood:

$$L_A = \frac{(\lambda_1 T)^4 e^{-\lambda_1 T}}{4!} \times \frac{(\lambda_A t_A)^0 e^{-\lambda_A t_A}}{0!} \frac{(\lambda_B t_B)^0 e^{-\lambda_B t_B}}{0!} \frac{(\lambda_C t_C)^0 e^{-\lambda_C t_C}}{0!} \frac{(\lambda_D t_D)^0 e^{-\lambda_D t_D}}{0!}$$

- Seems like overkill, but the degeneracies are important!

Log-Likelihood:

$$\begin{aligned}\ln L_A &= 4 \ln(\lambda_1 T) - 4\lambda_1 T - \ln(4!) \\ &\quad - \lambda_A t_A - \lambda_B t_B - \lambda_C t_C - \lambda_D t_D\end{aligned}$$

Optimized with $\lambda_A = \lambda_B = \lambda_C = \lambda_D = 0$, and then:

$$\lambda_1 = \frac{4}{T}$$

Substituting this all back into the original likelihood gives the "Profile" or "Concentrated" likelihood:

$$L_A = \frac{4^4 e^{-4}}{4!}$$

Chain Two:

Log-Likelihood

$$\begin{aligned} L_C = & \frac{(\lambda_1(t_4 + t_A))^1 e^{-\lambda_1(t_4+t_A)}}{1!} \frac{(\lambda_2(t_3 + t_B))^1 e^{-\lambda_B(t_3+t_B)}}{1!} \\ & \times \frac{(\lambda_3(t_2 + t_1 + t_E))^2 e^{-\lambda_3(t_2+t_1+t_E)}}{2!} \\ & \times \frac{(\lambda_C t_C)^0 e^{-\lambda_C t_C}}{0!} \frac{(\lambda_D t_D)^0 e^{-\lambda_D t_D}}{0!} \end{aligned}$$

Highlight: **fewer degenerate chains, and more active chains!**

Chain Two: profile/concentrated-likelihood:

$$L_C = \frac{1^1 e^{-1}}{1!} \frac{1^1 e^{-1}}{1!} \frac{2^2 e^{-2}}{2!} = \frac{2^2 e^{-4}}{2!}$$

Comparison of L_A and L_C is a race between $\frac{4^4}{4!}$ and $\frac{2^2}{2!}$.

Relative likelihood: $P(A|A \text{ or } C) = \frac{\frac{4^4}{4!}}{\frac{4^4}{4!} + \frac{2^2}{2!}} = .84$

Key Feature:

$$h(n) = \frac{n^n}{n!}$$

This kernel is *convex*. Breaking it up into smaller chunks reduces likelihood. That is:

$$h(n) > h(n - k)h(k)$$

Simpler explanations mean longer chains mean higher probability

Continuing on...

- Probabilistic explanation of Occam's Razor in a way that captures Dyen's notion of simplicity.
- How can these ideas be related to a notion of distance or divergence?
- How can divergence and probability be tied together formally, as the AAH posits?

A Start:

- With each location/culture k , there are a family of possible migratory histories \mathcal{H}_k that explain the underlying geography of the phylogeny.
- For $H_k \in \mathcal{H}_k$, define $N(H_k)$ as a count of the number of (non-degenerate) migratory chains in the history.
- Define $n(C)$ as the number of events in a non-degenerate migratory chain, and then define:
- $n_{H_k}^* = \max_{C_{ik} \in H_k} [n(C_{1k}), n(C_{2k}), \dots, n(C_{N(H_k)k})]$ - The maximum event count for the chains in H_k .

Definition: Dyen Divergence

Start with a function $D_{H_k} = m(n_{H_k}^*, N(H_k))$, where m is increasing in its first argument, and decreasing in the second. Define now the *Dyen Divergence* as

$$D_k = \max[D_{H_{1k}}, D_{H_{2k}}, \dots, D_{H_{Ik}}]$$

A family of divergence measures. Examples:

- $D_k^1 = n_{H_k}^* - N(H_k)$
- $D_k^2 = \frac{n_{H_k}^*}{N(H_k)}$

Age-Area Theorem

Suppose model assumptions hold, and define a Dyen Divergence measure. Then:

$$D_k \geq D_j \implies L_k \geq L_j$$

Further

$$\begin{aligned} k &= \arg \max [D_1, D_2, D_3, \dots, D_n] \\ \implies k &= \arg \max [L_1, L_2, L_3, \dots, L_n] \end{aligned}$$

Proof (sketch)

Note likelihood obeys

$$L_k \propto \prod_{j=1}^{N(H_k^*)} h(n_j), \quad \sum n_j = I$$

Because of convexity of $h(n)$, pile up as many n 's in as few chains as possible. Analogy: a risk-loving investor with fixed assets and a bunch of investment choices.

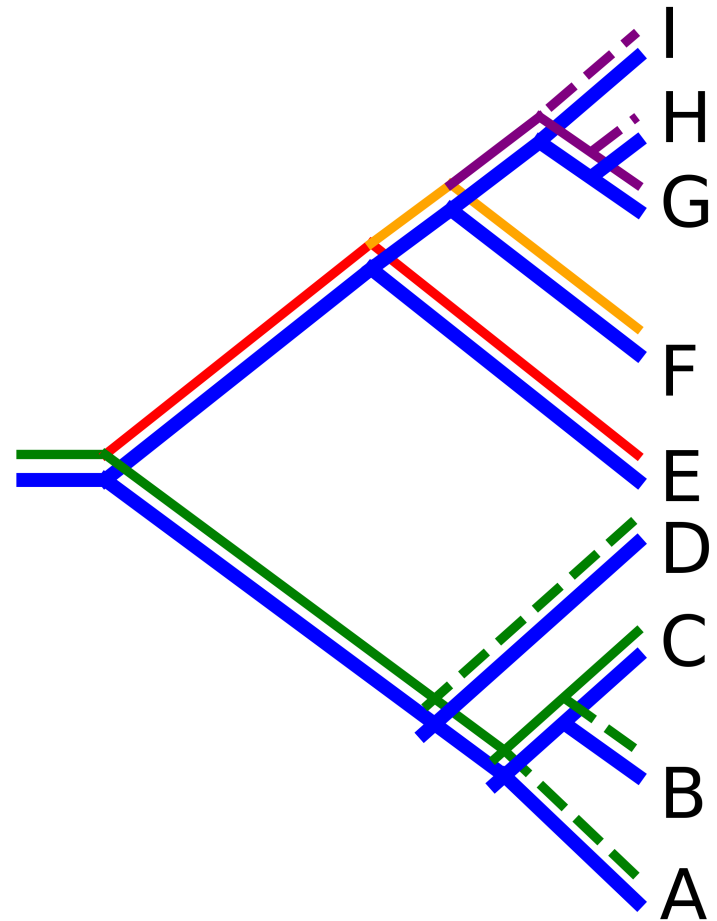
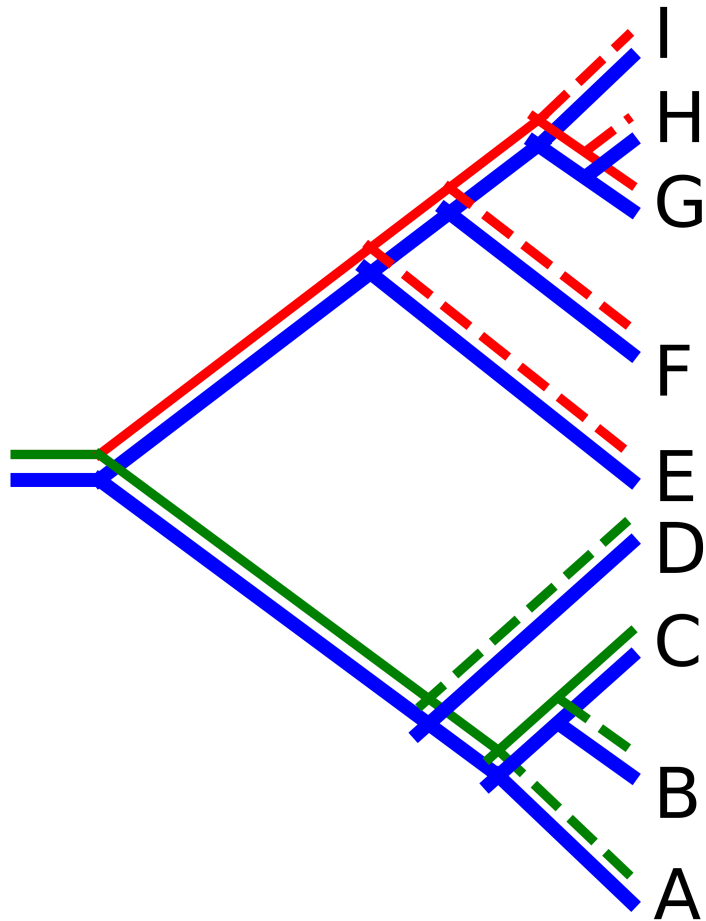
Also, finite ordered sets have a maximum.

Illustration of Dyen Divergences:

$$D_i^1 = \frac{n_{H_k^*}^*}{N(H_k^*)}$$

$$D_i^2 = n_{H_k^*}^* - N(H_k^*)$$

Additional Example: E versus I



Comparison of divergence measures

- If E is the point of origin:
 - chain from E to D to A to B to C
 - chain from E to F to I to G to H
 - Dyen Divergence - 2 chains, 4 events each. $D^1 = 2$, $D^2 = 2$.
- If I is the point of origin:
 - chain from I to D to A to B to C
 - chain from I to E, chain from I to F
 - chain from I to H to G.
 - Dyen Divergence - 4 chains, with 4, 1, 1, and 2 events. $D^1 = 0$, $D^2 = 1$.

Likelihoods

- For E, calculating it out gives relative likelihood as .84.
- A better contender to E, however, is D. Two chains, one with 5 events, and one with 3.
 - Dyen Divergence - $D^1 = 3$, $D^2 = 2.5$
- Oftentimes, this isn't obvious from looking at the Tree...

Bells and Whistles

- Known branch lengths -
 - Poisson becomes Exponential distribution
 - Important case for including migration model with tree estimation
- Algorithm for calculating probabilities/divergences -
 - One can traverse the tree backwards, using dynamic programming to pick out the most likely continuation path
- Include other information in the decision
 - Physical distance
 - Prior knowledge of location, time, or split time

Micro foundations

- Why would one believe the exponential/Poisson arrival rate story?
- Idea: stochastic population growth model
 - Chain propelled by positive resource/technology shock.
 - If the population in a location hits a barrier, the positive shock instantaneously dissipates.
 - Population splits according to an arbitrage condition, and a fraction moves on to a new area.
 - A new shock parameter is drawn in the location.

Point of emphasis

It is important that the shock dissipates, and does not instead reset. Otherwise, locations would be observed to continually expel migratory groups, and mass migrations would occur from all locations continually, like a branching process.

Formal approach (Baker, 2008):

- Utility, children, and (net) income are equal (at a location):
- Income has a fixed component, a congestion component, and a stochastic component: $y_t = 1 + r(1 - \frac{p_t}{K}) + \sigma(\epsilon_{t+\Delta} - \epsilon_t)$
- Total population next period, $p_{t+\Delta}$ is $p_t k_t$ or:

$$p_{t+\Delta} = p_t y_t = p_t \left[1 + r \left(1 - \frac{p_t}{K} \right) + \sigma(\epsilon_{t+\Delta} - \epsilon_t) \right]$$

As Δ gets small...

$$dp = rp \left(1 - \frac{p}{K}\right) + \sigma p dz$$

Stochastic Logistic population growth model: drift $rp \left(1 - \frac{p}{K}\right)$ and variance $\sigma^2 p^2$.

Crucial property: *stationary distribution*.

Theorem

Nobile, Ricciardi & Sacerdote (1985)

If a sde has a stationary distribution, the first passage time to a barrier B given initial population p_0 , is **approximately exponential**:

$$g(B, t|p_0) \approx \frac{1}{t_1(B|p_0)} e^{-\frac{t}{t_1(B|p_0)}}$$

where t_1 is the first moment of the distribution of the first passage time distribution.

Rounding out the Story

1. Moving to a new location involves a cost c and requires a minimum population M to move.
2. If $p = B$ is achieved, K falls immediately to $K - D$ for the current generation.
3. Current generation: Splits to avoid the precipitous drop in utility.
4. Upon exit, a new B, K, D combination is drawn.

Parameterization

- Utility is $1 + r(1 - \frac{p_t}{K})$.
- $B = K$.
- Costs of moving are $c = r$, so utility in a new location is maximally 1, while utility before the barrier is hit in the original always greater than one.
- When $p = B = K$, arbitrage condition dictates size of staying population and emigrating population:

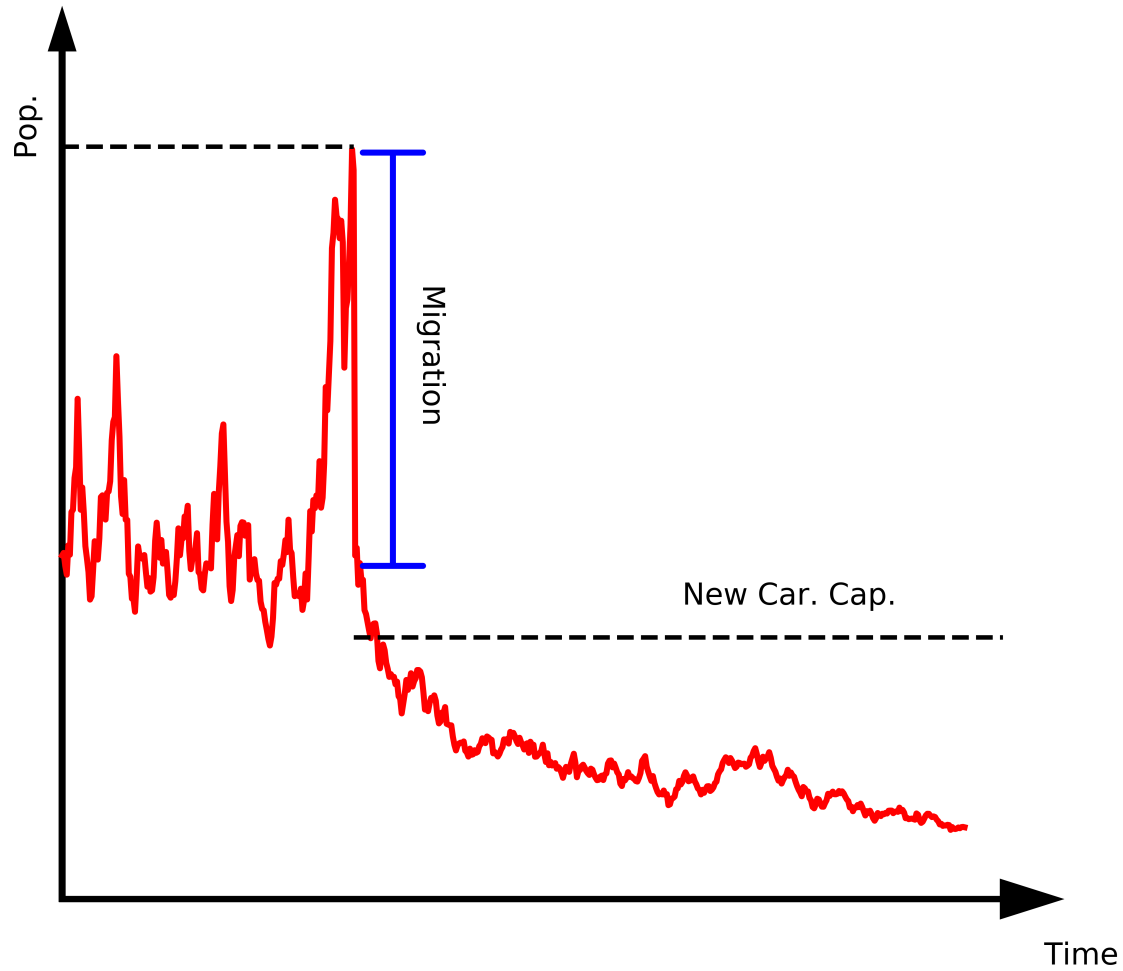
Arbitrage

If m is the emigrating group, then:

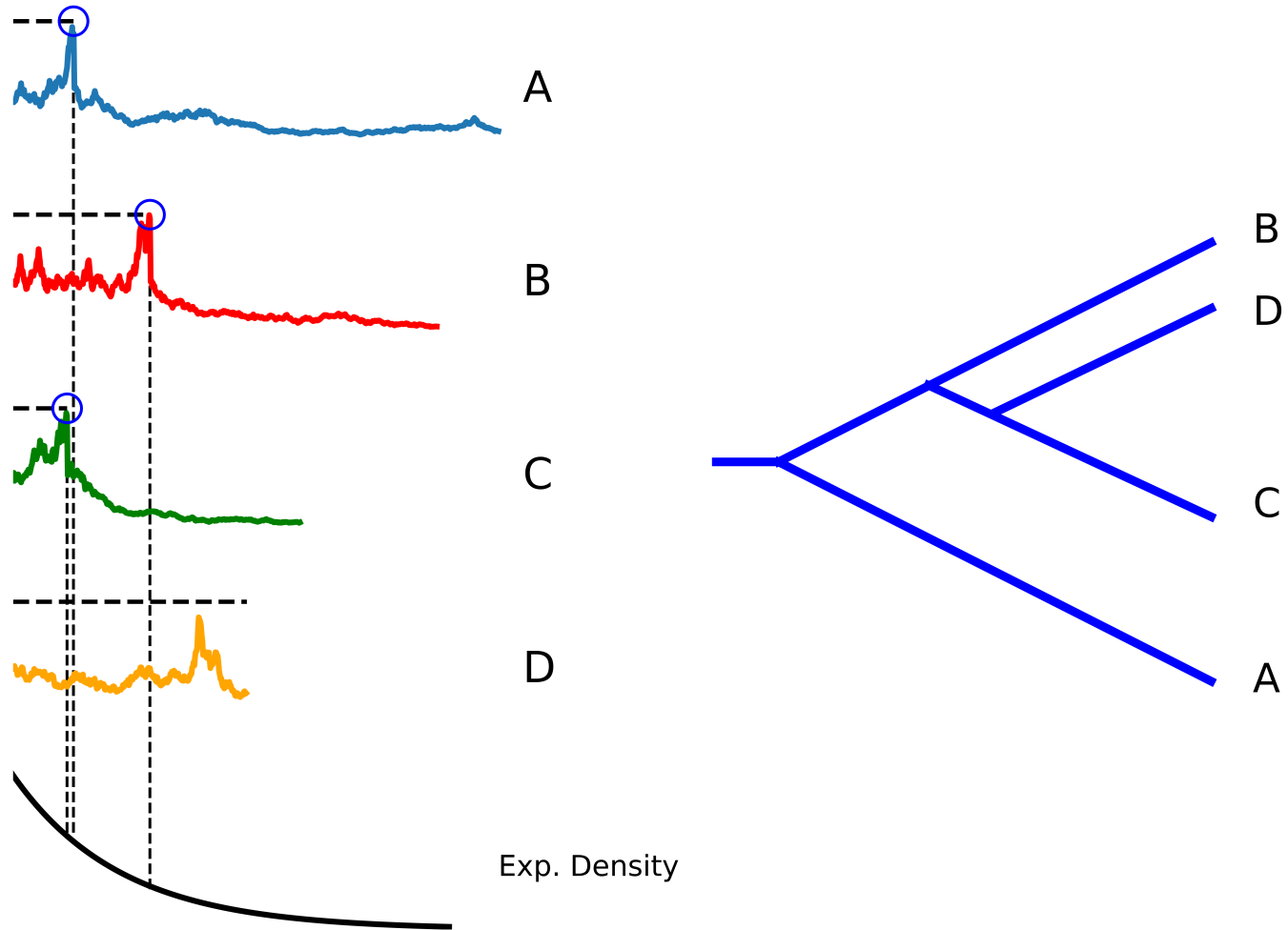
$$1 + r \left(1 - \frac{B - m}{K - D} \right) = 1 - c + r \left(1 - \frac{m}{K} \right)$$

Results in $B = K, p_0 = m = \frac{DK}{2K - D}$

Illustration in which new values of B , K , D are drawn so that the location is effectively dormant.



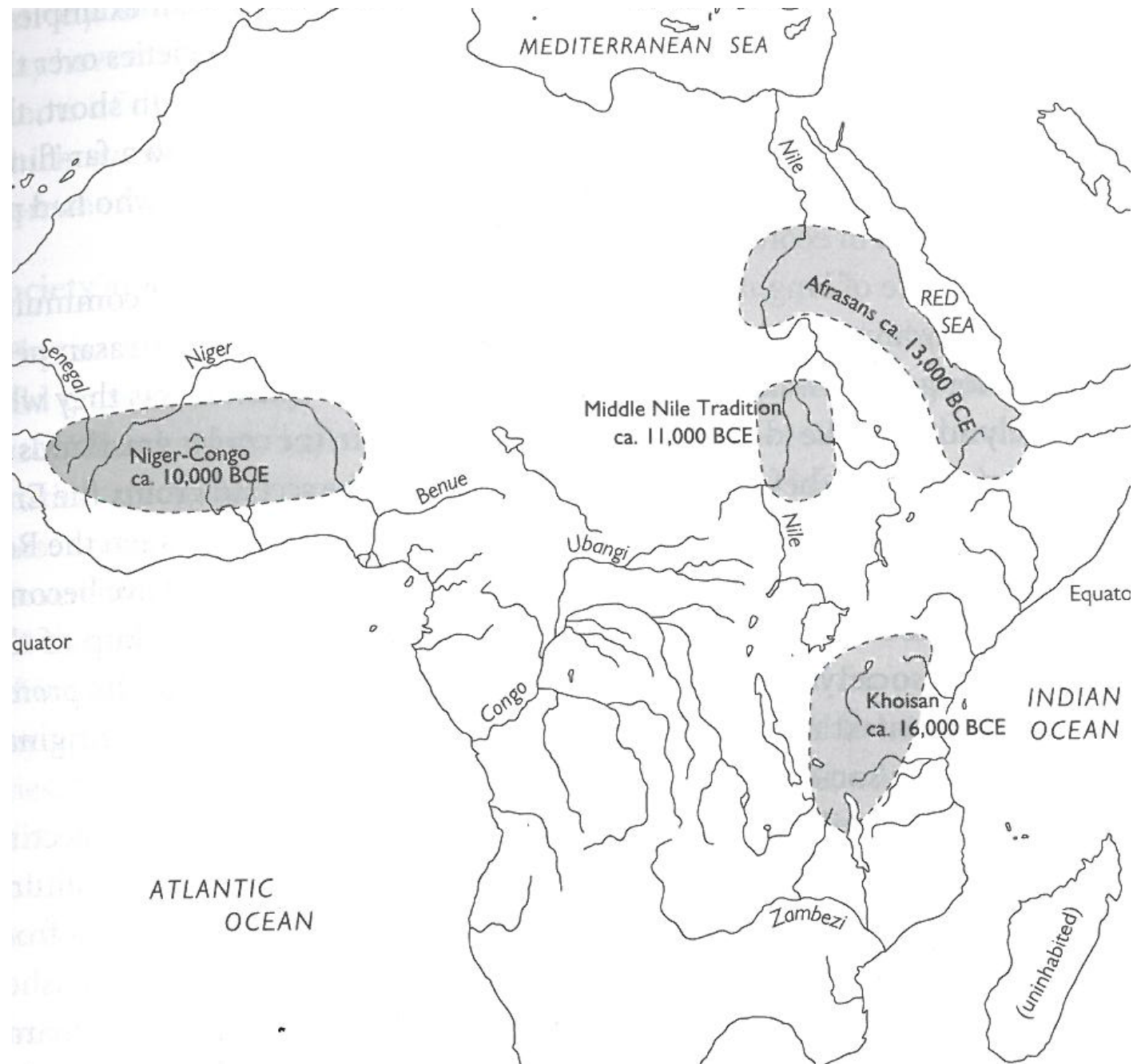
Further illustration:



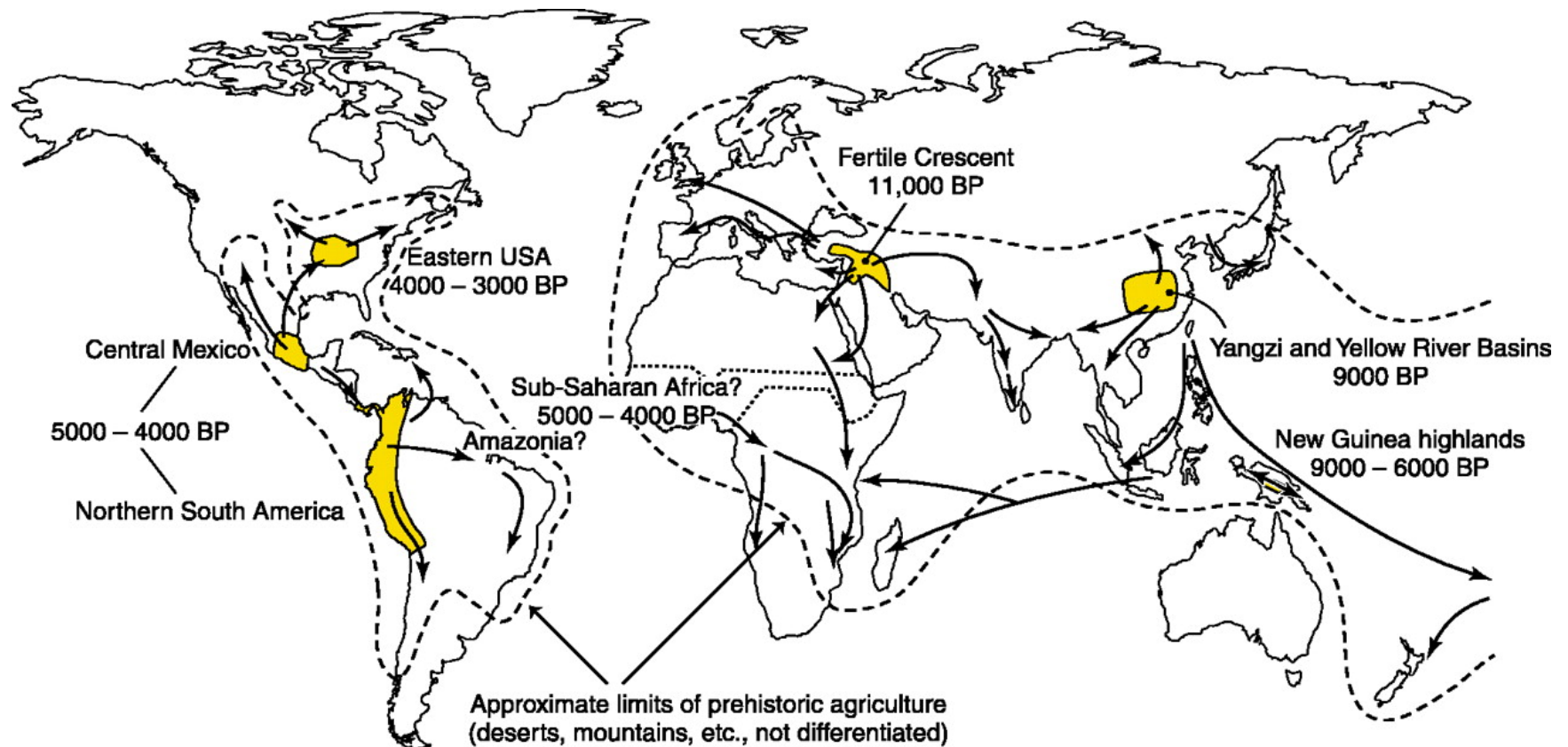
Applications

- Afrasan or Afroasiatic and its point of origin. Arabic and Semitic languages, Ancient Egyptian, and Ethiopiac languages as well. Where did this Phylogeny originate?
- NaDene phylogeny and its point of origin. Simulating *spatial and temporal points of origin*.

From Ehret (2000) - origins of Afrasans



Diamond and Bellwood (2003)- origins



Probabilities of Origin Points

- These are "Tree Ring" type maps!
- [Link to Afrasan Map](#)
- [Link to Na Dene Map](#)
- [Link to Khoisan Map](#)

Sampling NaDene History

- Idea: blend known branch lengths, migratory distances, and estimation of a linguistic phylogeny using standard methods:
 - Lexicostatistics/glottochronology
 - Bayesian Priors on some dates, MCMC sampling (Baker, 2015)
- Create a probability distribution over histories.
- Idea follows Baxter and Ramer (1993), Mace and Pagel (1993), and Dogolpolsky (1985) - use the type of first letter for Swadesh lists.

Swadesh Lists

Essentially, a list of words that are:

- Very resistant to borrowing
- Evolve very slowly somewhat like a genetic code.

Words: I, you, we, one, two, person, fish, dog, louse, tree, leaf, skin, blood, bone, horn, ear, eye, nose, tooth, tongue, knee, hand, breast, liver, drink, see, hear, die, come, sun, star, water, stone, fire, path, mountain, night, full, new, name.

The Na-Dene letters/Dogolpolsky classes

cwittl-cwctssscwllclttyNyccici ccsts-s	HAIDA
ci tll yycl tttctcl Ncl icssi csycccccti ctci	CHIRICAHUA
sNNcl tll sccsttcsssssssssi i i i ssttctttcs	NAVAJO
sNt Nt l yct-tctstnc-cl c- Nl cc-sytsct-l mi	JICARILLA
sNN NN l yc--sctcNl wsc l -- Nl ct-sstscct-l p-y	JICARILLA APACHE
sNN Ncl Nyct-cctcNNcsc l c- Nl ct-cctcct-cp-c	SAN CARLOS
wNN Ncl l yit-cctcNNc-cl c- Nl cc-sytsct-cmi	LIPAN
sNN itl l yii-tctcNcccc l c- Nl cc-cstcci-cc-i	BEAVER
sNN Nt l ytt-tttcNctcNt-titt-stttct-cp-i	CARRIER
sNN Nt l yti-iiiiiiiiii l t-i i cp-sptccN-ti-i	C CARRIER
sNN Nt l sti-ttiiiiiiii cNt-citt-sstccN-ti-i	KUTCHIN
iiii Nt l -tiiiiiiiiiiii i i i i -tt-sstcctci t-p	HARE
sNci Nt l yct-tttcNnci cl t- Nl cc-ssttct-cp-i	CHIPEWYAN
sNN itl l ytt-tctctNwc- N- ti cc-sstcct-c--y	SARCEE
sNN Nt l Nyct-tctcNl csc l p- l i cc-ysccci -l p-i	HUPA 2
sNN NN l yct-tctcNccsc l p- l i cc-yctcci -l p-i	MATTOLE
sttt Nl t-t--tityttwsc l pstc-c--sc-ci ctc-s	KATO
sNN Nt l yNt-tctcNccsc l p- l i cc-yctcci -l p-i	GALICE
sNN Nt l yci itci i i i i i i i i i yttysstcci ttii i	TANACROSS
sNN NN cyct-tctcNl csc l p- l i cc-cstcci -l p-i	EYAK
ittssNcc-ccccsNcccctcsttNccci cccccctsscc	TLINGIT

Estimation using MCMC methods

- We want density $P(H, T)$.
- So, simulation by drawing from conditional densities $P(H|T)P(T)$ and then $P(T|H)P(H)$
- Can add in prior on certain split dates and on tree structure.
- Simulation from distribution estimated using linguistics
- Backup

In Conclusion:

- Recent literature on diversity is getting more sophisticated and multidisciplinary
- Doesn't mean it should sacrifice rigor
- An effort to formalize and operationalize some of this
- **In the future:** formal models of borrowing, interaction, and cultural evolution.

