Foundations of the Age-Area Hypothesis

Matt Baker

Background

- Research agenda: the economic basis for indigenous institutions
- Big question: How environment, technology, and institutions co-evolve

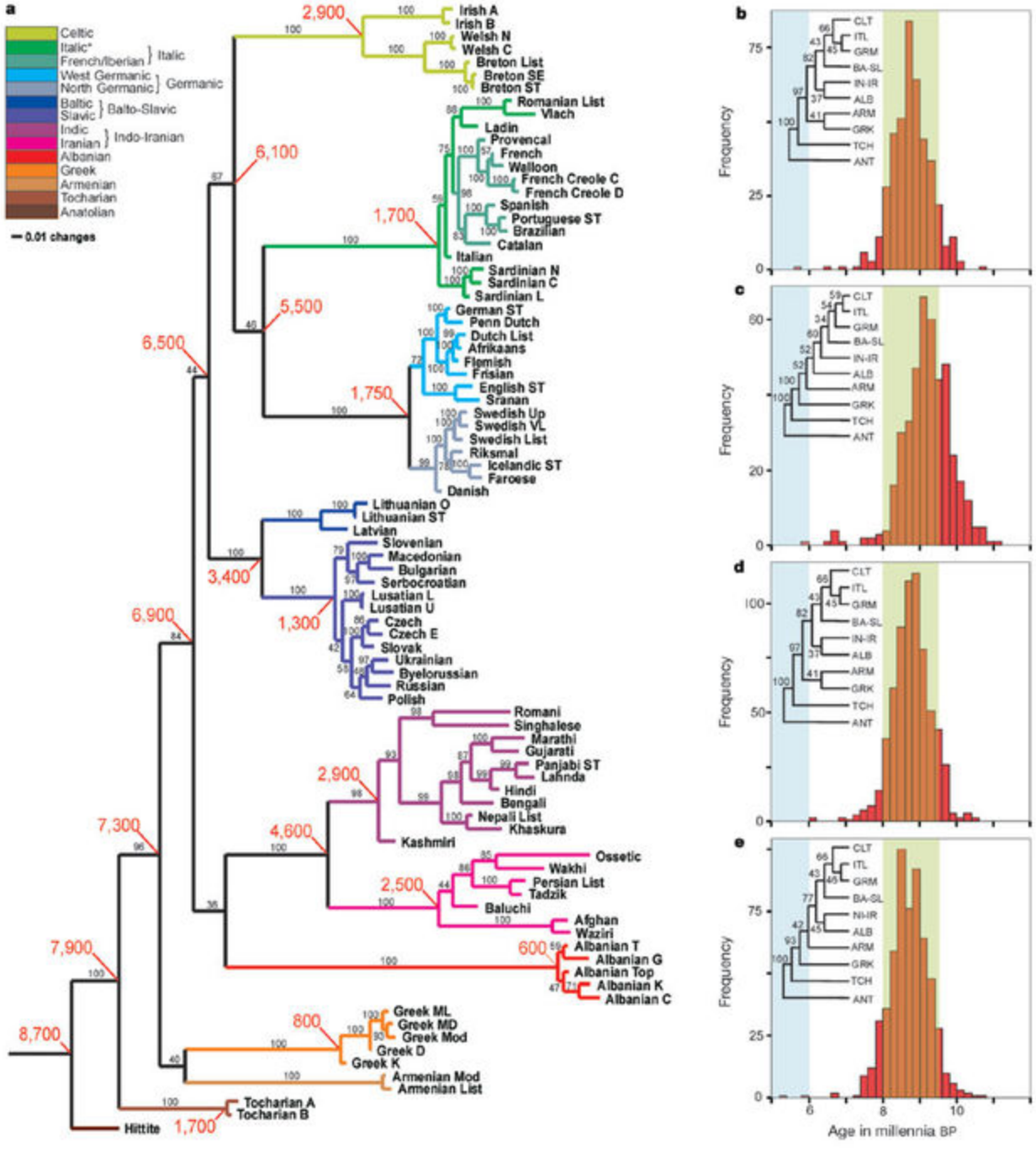
Recently

- Applications in economic growth
- Computational linguistics, computing power
- Incorporation of geographical data into analyses.

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

Cultural similarities

- Closely related to genetic similarities
- Computational linguistics - treat aspects of language like a genetic code with drift.
- Build *Phylogenies* of related cultures; epitome Mace (2006).
- Atkinson and Gray (2006) example: Indo-European Tree.
- Fairly sophisticated machinery for doing this!


a
Celtic
Italic*
French/Iberian
Italic
West Germanic
North Germanic
Germanic
Baltic
Slavic
Balto-Slavic
Indic
Iranian
Indo-Iranian
Albanian
Greek
Armenian
Tocharian
Anatolian
0.01 changes
2,900
6,100
1,700
5,500
6,500
1,750
3,400
1,300
6,900
2,900
7,300
4,600
2,500
7,900
600
800
8,700
1,700
Irish A
Irish B
Welsh N
Welsh C
Breton List
Breton SE
Breton ST
Romanian List
Vlach
Ladin
Provencal
French
Walloon
French Creole C
French Creole D
Spanish
Portuguese ST
Brazilian
Catalan
Italian
Sardinian N
Sardinian C
Sardinian L
German ST
Penn Dutch
Dutch List
Afrikaans
Flemish
Frisian
English ST
Sranan
Swedish Up
Swedish VL
Swedish List
Riksmal
Icelandic ST
Faroese
Danish
Lithuanian O
Lithuanian ST
Latvian
Slovenian
Macedonian
Bulgarian
Serbocroatian
Lusatian L
Lusatian U
Czech
Czech E
Slovak
Ukrainian
Byelorussian
Russian
Polish
Romani
Singhalese
Marathi
Gujarati
Panjabi ST
Lahnda
Hindi
Bengali
Nepali List
Khaskura
Kashmiri
Ossetic
Wakhi
Persian List
Tadzik
Baluchi
Afghan
Waziri
Albanian T
Albanian G
Albanian Top
Albanian K
Albanian C
Greek ML
Greek MD
Greek Mod
Greek D
Greek K
Armenian Mod
Armenian List
Tocharian A
Tocharian B
Hittite
b
c
d
e
CLT
ITL
GRM
BA-SL
IN-IR
ALB
ARM
GRK
TCH
ANT
Frequency
Age in millennia BP

Questions:

- Where did this tree originate?
- How did the peoples of the tree come to be where they are?
- Which related cultures have been in close proximity, and for how long?

Questions of geography, cultural/lingustic drift, and time.

The Age-Area Hypothesis (AAH)

- Sapir (1916) - the root of the tree is the geographical point of origin.
- Recursive application - migratory routes
- Used to resolve historical debates, but also could be important in creating new theories

Old applications and continuing debates

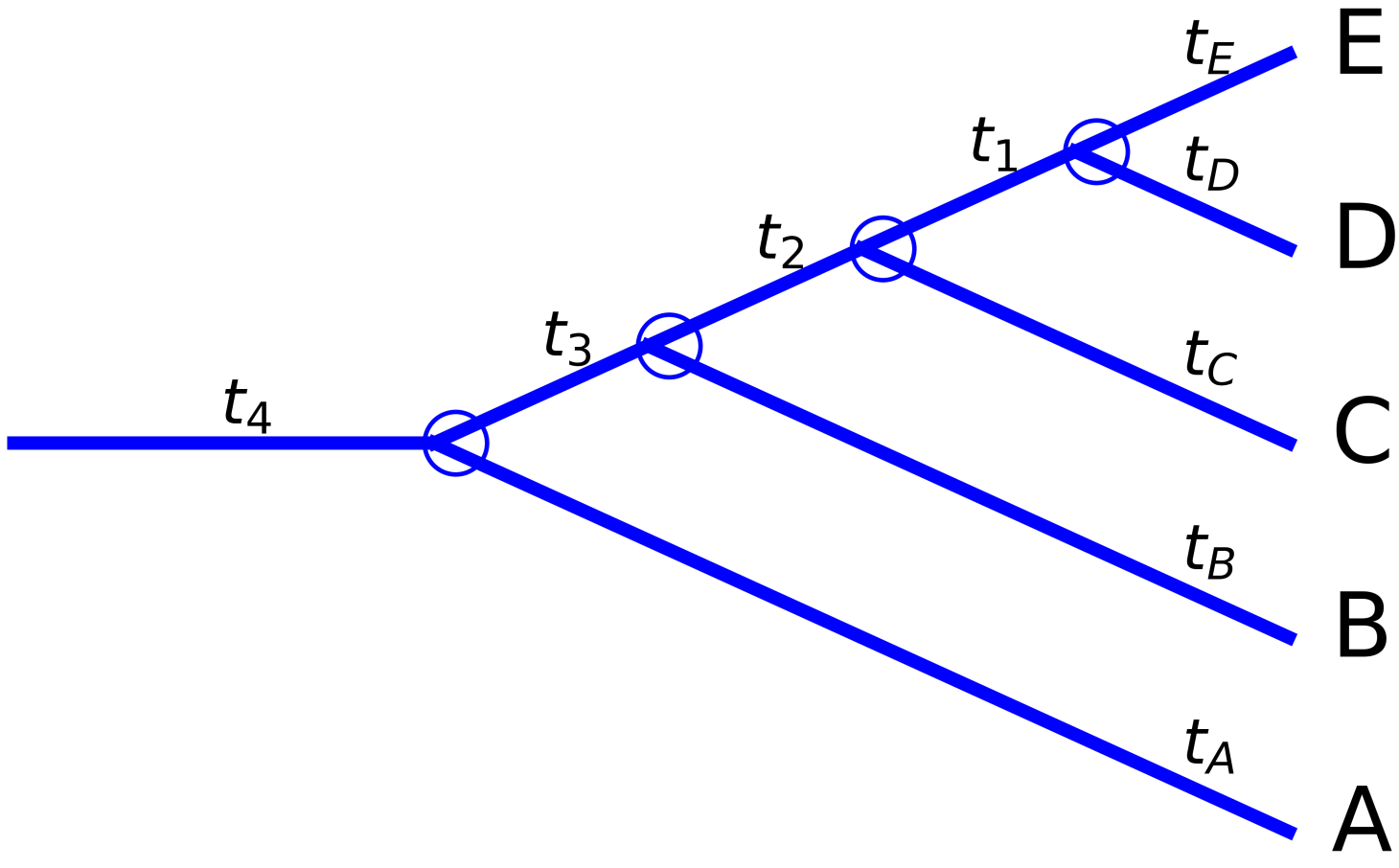
- Origins of Athabaskan/Na-Dene speakers
- Indo-European origins
- Afro-Asiatic origins
- Spread of Bantu peoples
- Native American population dispersal

On the need or theory...

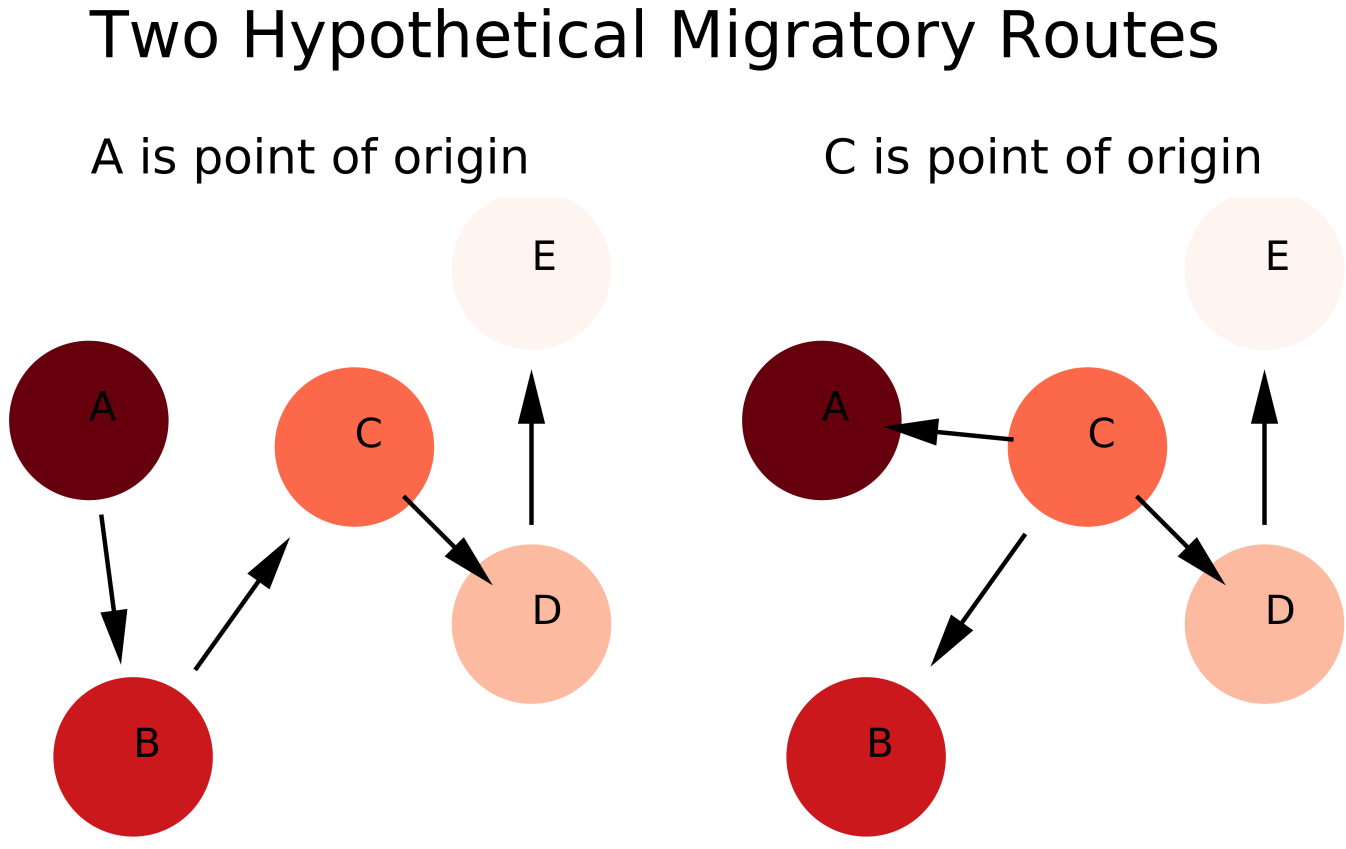
Greenhill and Gray (2005) write: "many expansion scenarios are little more than plausible narratives. A common feature of these narratives is the assertion that a particular lineof evidence (archaeological, linguistic, or genetic) is 'consistent with' the scenario. 'Consistent with' covers a multitude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

So why believe the AAH (or not)?

- Occam's Razor?
- Minimum effort or # of moves?
- Dyer (1956, p. 613) seems to hit 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."


t_E
E
t_1
t_D
D
t_2
t_3
t_C
C
t_4
t_B
B
t_A
A

Problem Preview



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 observe drift. Latter seems more complex. Howso?
- Note "minimum moves" doesn't get us very far. Both have four moves!

Basic Model:

- **Assume a full, rooted binary tree**
    - Tree with z terminal nodes will have $z - 1$ internal nodes, which are the minimal number of 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

Migratory Chain

A sequence of "forward moving" migratory events that end at a terminal node/taxa/culture.

Migratory History

A collection of Chains spanning the whole tree, with a "deepest chain" starting at a given location.

Observations:

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

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

A family of divergence measures. Examples:

- $D^1_k = n^*_{H_k} - N(H_k)$
- $D^2_k = \frac{n^*_{H_k}}{N(H_k)}$

Distributions over Histories

- Simulation

HOw about this brah?