

Language diversity and its relationships with genes, history and space

A rather not very coherent amalgam of interesting bits

Dan Dediu

The Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

Dan.Dediu@mpi.nl

Biological Evolution and Diversification of Languages (BEDLAN)

University of Turku

23rd September 2010



Overview

- Genes and Language:
 - Evolutionary, population and individual levels
 - **Genetic biasing of language**
- Language through time:
 - **Phylogenetic stability of typological features**
- Dialect variation
 - English dialect dynamics
 - **Language shift in Early Modern Dunkirk**



Genes and Language

The Species/Evolutionary Level

- Language & speech rest on “**species-specific**” genetic foundations
 - Not necessarily language-specific
 - Building on **existing structures & behaviors**
 - Probably **subtle changes**
 - Mostly in regulatory regions



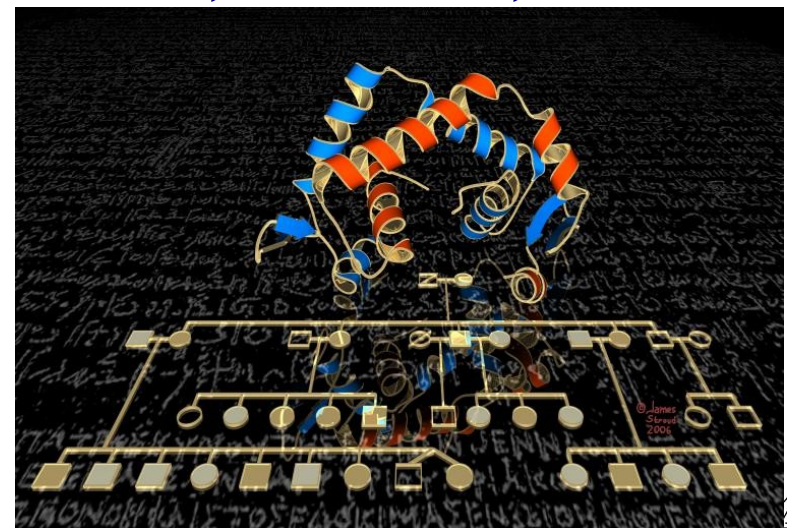
Human vocal tract
www.llas.ac.uk



Genes and Language

The Individual/Ontogenetic Level

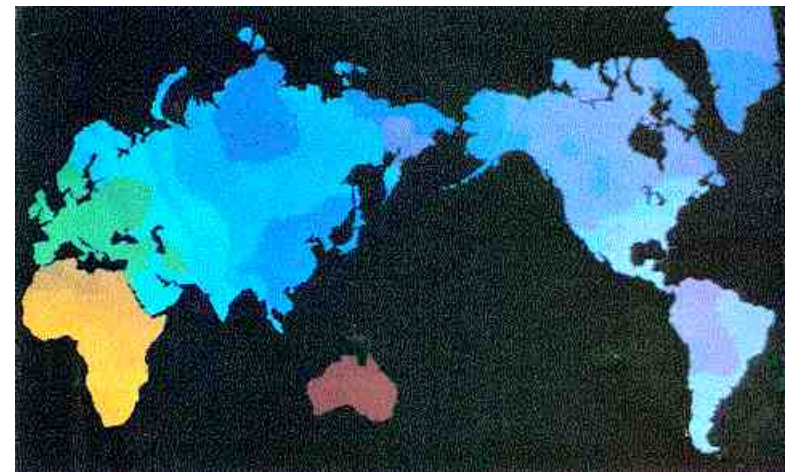
- Individual differences in speech & language have a **strong genetic component**
 - Normal & pathologic
 - **Heritability** and specific **genes**
 - *FOXP2, CNTNAP2, KIAA0319, ROBO1, ...*



Genes and Language

The Population/Glossogenetic Level

- Languages and genes as “**travel companions**” through space & time
 - They are changing together
 - **Demographic processes** affect both in parallel ways
 - Produce **accidental correlations** between language relatedness and genetic similarity



Genetic gradients across the world
The History and Geography of Human Genes



Genetic Biasing of Language

The Idea

Language is **shaped by** the

- **Production** and **perception** systems
 - **Brain/cognition**
 - **Social & cultural** factors and processes
- some have *genetic components*

Adaptation to these through **cultural evolution**

⇒ *genetic biases can affect language*

But no “gene for Chinese”!



Genetic Biasing of Language

Modeling and Experimental Support

- Models of **language evolution**: *small biases have a complex dynamics when culturally transmitted* (Kirby et al., 2007; Dediu, 2009)
- Models of **cultural evolution**: “*Content*”/“*Direct*” *Biases* (Boyd & Richerson, 1985, 2005); *Niche Construction* (Laland, Odling-Smee, Feldman, 2003)
- Experimental amplification of biases through *cultural chains*: compositionality, systematicity, phonology, etc. (Smith & Wonnacott, 2010; Kirby et al., 2007, etc.)



Genetic Biasing of Language

Two Early Suggestions

- **Darlington** (1947,1955) & **Brosnahan** (1961):
 - Europe: correlation *blood groups* and *phonetics* (interdental fricatives, front rounded vowels)
 - Dismissed at the time (e.g., Swadesh, 1961)
- **Italian and Yoruba vowels** (Ladefoged, 1984; Disner, 1983)
 - 7-vowel systems (/i e ε a ɔ o u/), **F₂ lower** in Italian

“This does not, of course, imply that a Yoruba could not learn perfect Italian. *Any individual speaker could compensate* for the overall, statistical, difference in headshape” (Ladefoged 1984: 86)

 - Genetic biasing through *vocal tract anatomy*



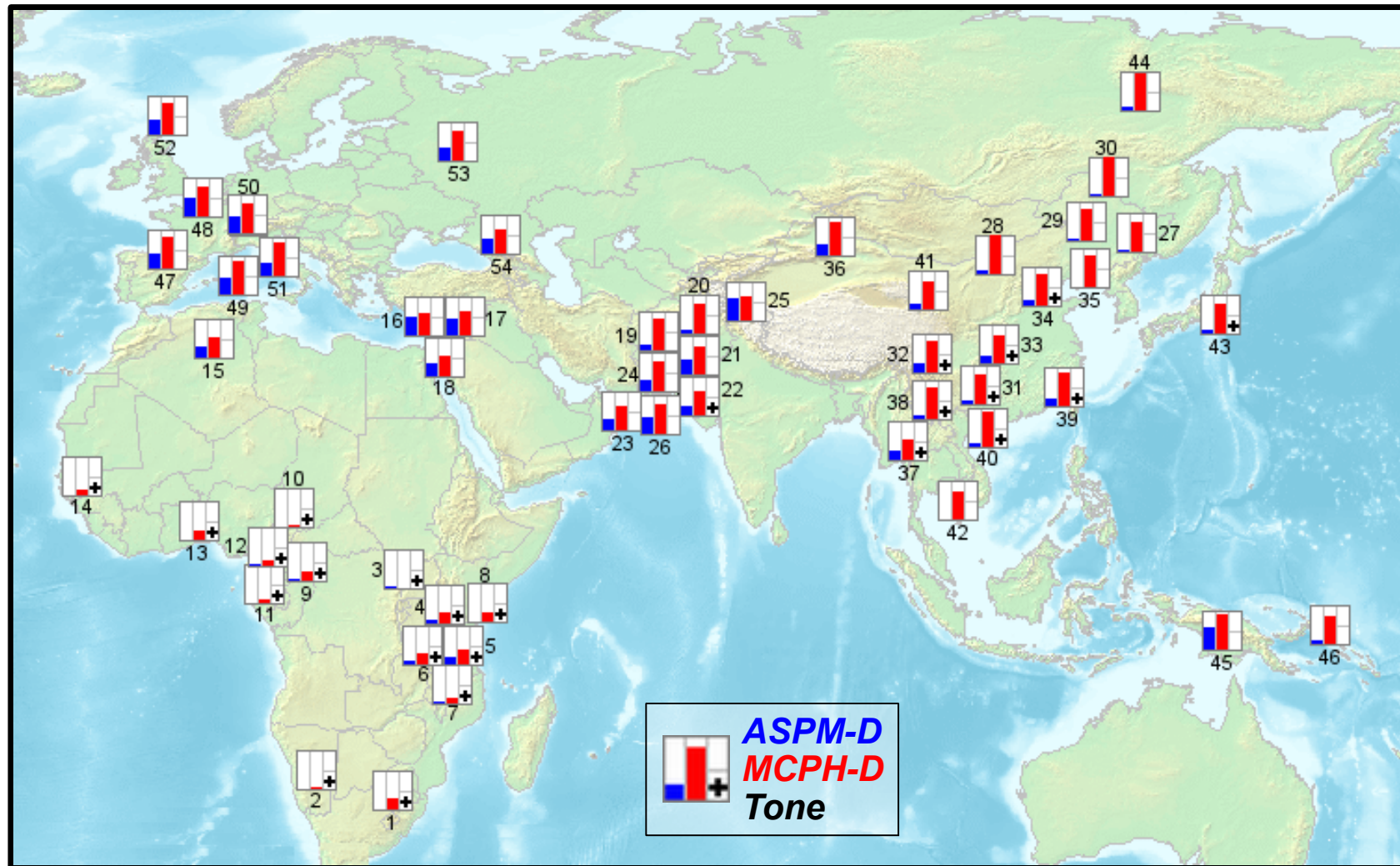
Tone, *ASPM* & *Microcephalin*

Two Genes Involved in Brain Development

- ***ASPM*** (1q31), ***Microcephalin*** (8p23)
- Recessive primary microcephaly
- Probably affects the *number of asymmetric divisions*
- *Accelerated evolution* in the lineage leading to humans
- **Derived haplogroups** (*ASPM-D*, *MCPH-D*):
 - ***Recent*** origin
 - Marked *geographic structure*
 - Recent/ongoing *natural selection* (???)
 - Unknown phenotypic effects



Tone, ASPM & Microcephalin



1. SE-SW Bantu, 2. San, 3. Mbuti, 4. Masai, 5. Sandawe, 6. Burunge, 7. Turu, 8. NE Bantu, 9. Biaka, 10. Zime], 11. Bakola, 12. Bamoun, 13. Yoruba, 14. Mandenka, 15. Mozabite, 16. Druze, 17. Palestinian, 18. Bedouin, 19. Hazara, 20. Balochi, 21. Pathan, 22. Burusho, 23. Makrani, 24. Brahui, 25. Kalash, 26. Sindhi, 27. Hezhen, 28. Mongola, 29. Daur, 30. Orogen, 31. Miao zu, 32. Yizu, 33. Tujia, 34. Han, 35. Xibo, 36. Uyгур, 37. Dai, 38. Lahu, 39. She, 40. Naxi, 41. Tu, 42. Cambodian, 43. Japanese, 44. Yakut, 45. Papuan, 46. NAN Melanesian, 47. French Basque, 48. French, 49. Sardinian, 50. N Italian, 51. Tuscan, 52. Orcadian, 53. Russian, 54. Adygei.



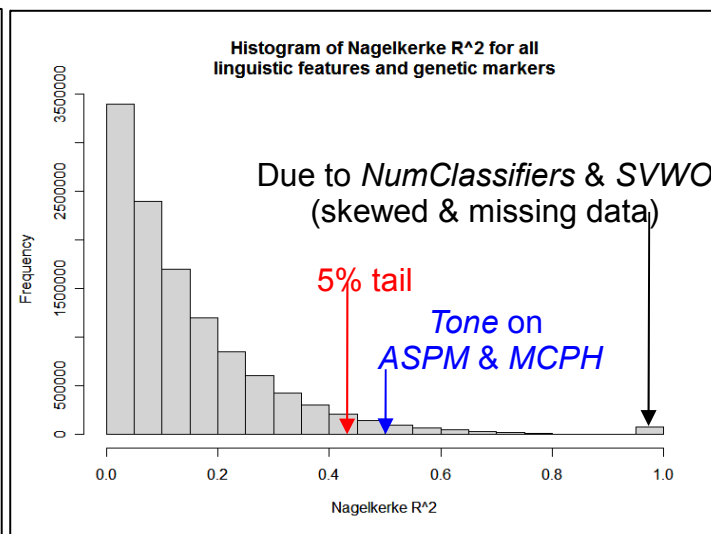
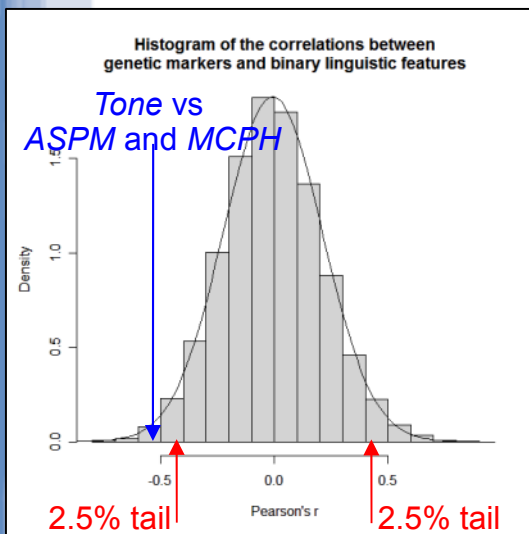
Tone, ASPM & Microcephalin

Database, Methods & Results

49 populations, 981 genetic markers, 26 binary features

- $r_{ASPM-Tone} = -0.53$, $r_{MCPH-Tone} = -0.54$, $p < 0.05$, top 1.4%
- **logistic regression:** $p < 0.05$, Nagelkerke's $R^2 = 0.53$, 73% corr. classif., in top 2.7%.

Mantel corr: *geography, genetics, typology & history:*

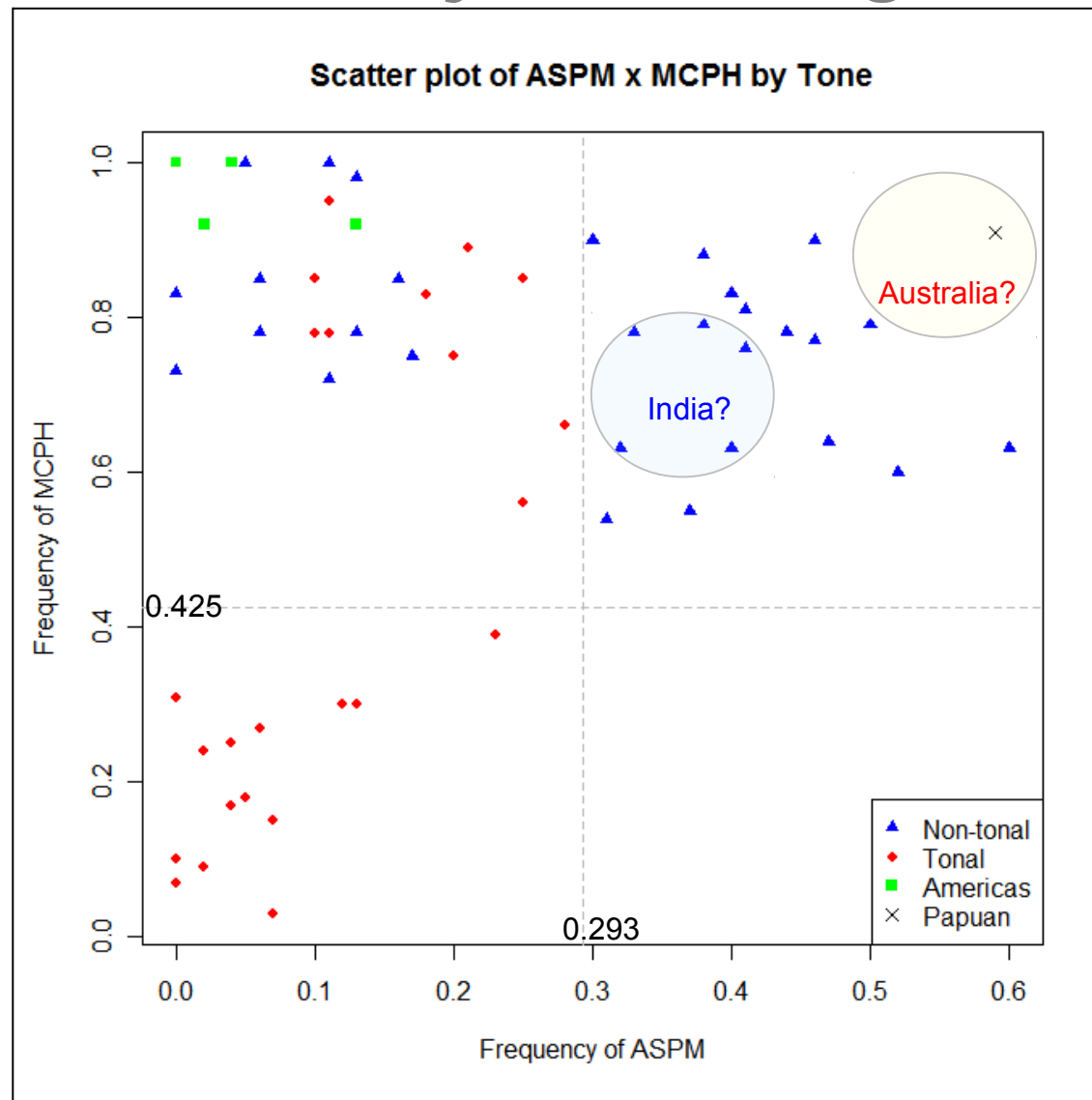


- controlling geography:
 $r = 0.291$, $p = 0.003$
- & controlling history:
 $r = 0.283$, $p = 0.000$



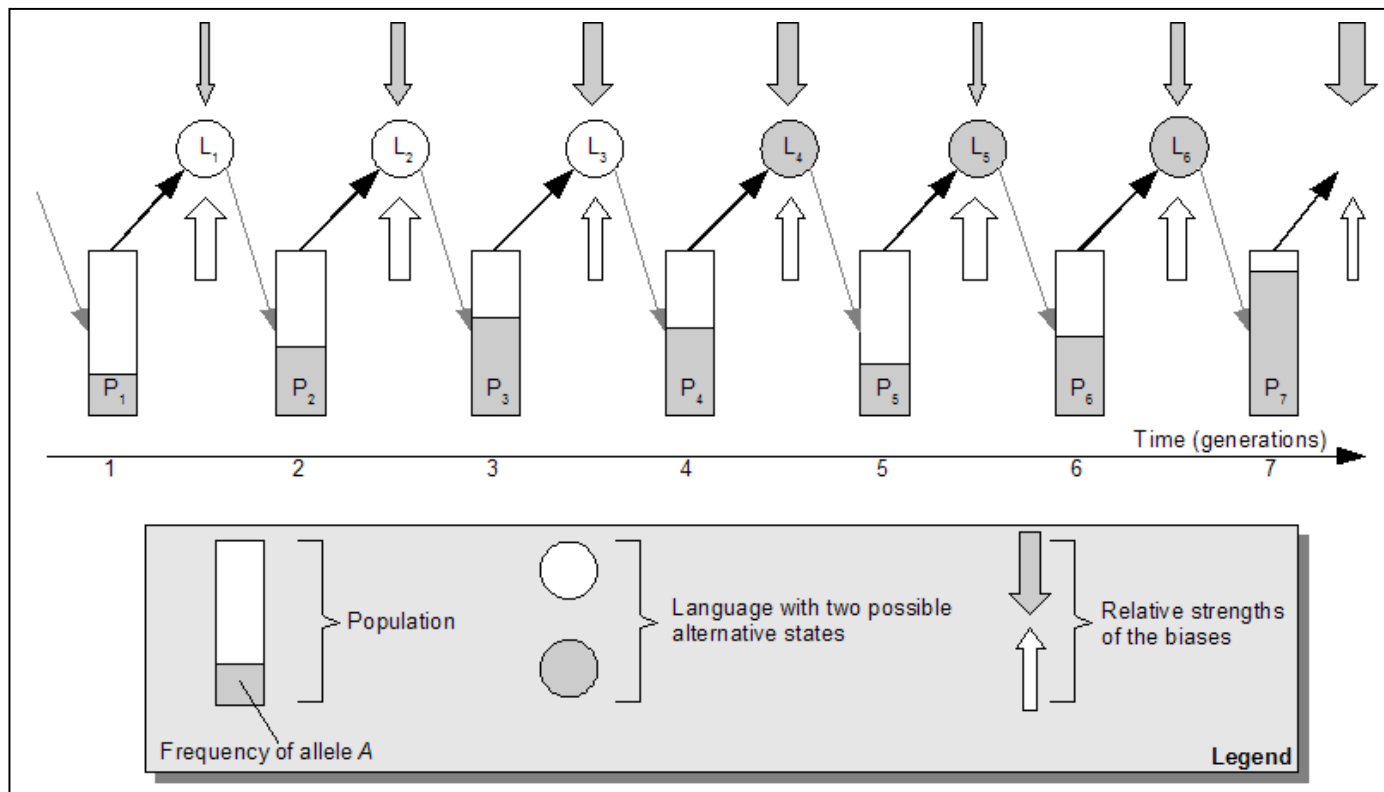
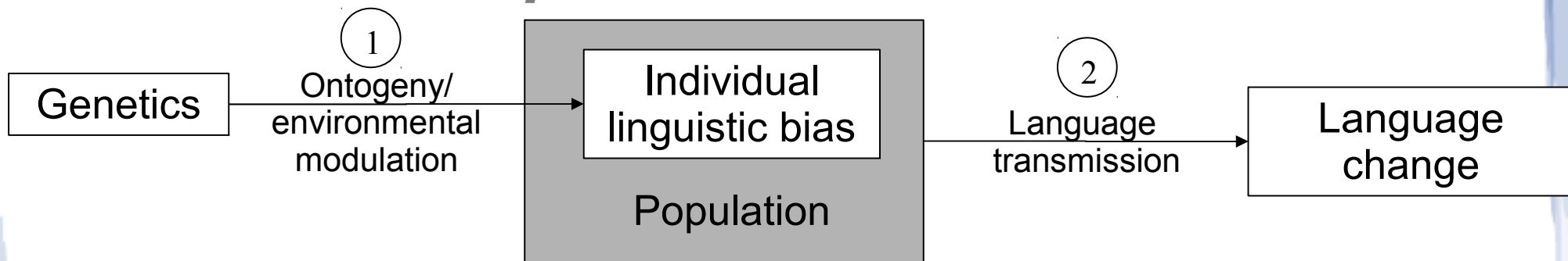
Tone, ASPM & Microcephalin

Summary of Findings



Tone, ASPM & Microcephalin

Proposed Mechanism



Tone, *ASPM* & *Microcephalin*

Other Supporting Evidence & Future Work

- Christiansen, Kelsey & Tomblin (pc. 2008): preliminary *association* between an *ASPM* SNP and several measures (probably) related to *phonology*
- Dediu (2010): tone seems to be **more stable** than expected
- Effects on *brain anatomy* (Rimol et al., 2010)
- **Work in progress:**
 - *Operationalization* of the bias & association with “ethnicity”, *ASPM* & *Microcephalin*
 - *Historical characteristics* of tone
 - More specific effects on brain structure



Phylogenetic stability

The idea

- **Vocabulary** (cognacy): can be *very stable* (Pagel et al 2007, Greenhill et al 2010)
- **Typology**: can also be very stable (Hunley et al 2008) but *differs across language families* (Greenhill et al 2010)
- What about **tendencies**? → small number of lg. fams
- The genetic biasing hypothesis → *tone should be stable*



Phylogenetic stability

Principles & Methods

- **Bayesian phylogenetics** (2 software packages: MrBayes 3 & BayesLang) → rate estimation
- As many **lg.fams** as possible (**41** in total) from **2** classifications (Ethnology & WALS)
- As many **typological features** as possible (**68** polymorphic & **86** binary; WALS)
- As many **outgroups** as possible (**23** language isolates)
- Estimate rates → rank features → compare ranks across software packages, data codings, outgroups,...



Phylogenetic stability

Results

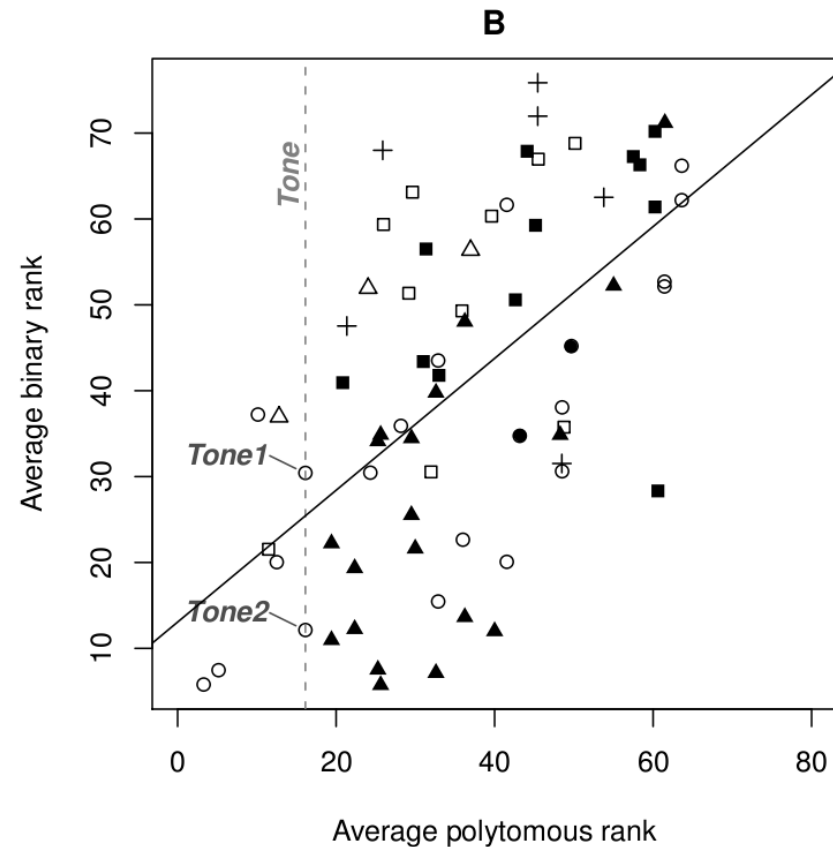
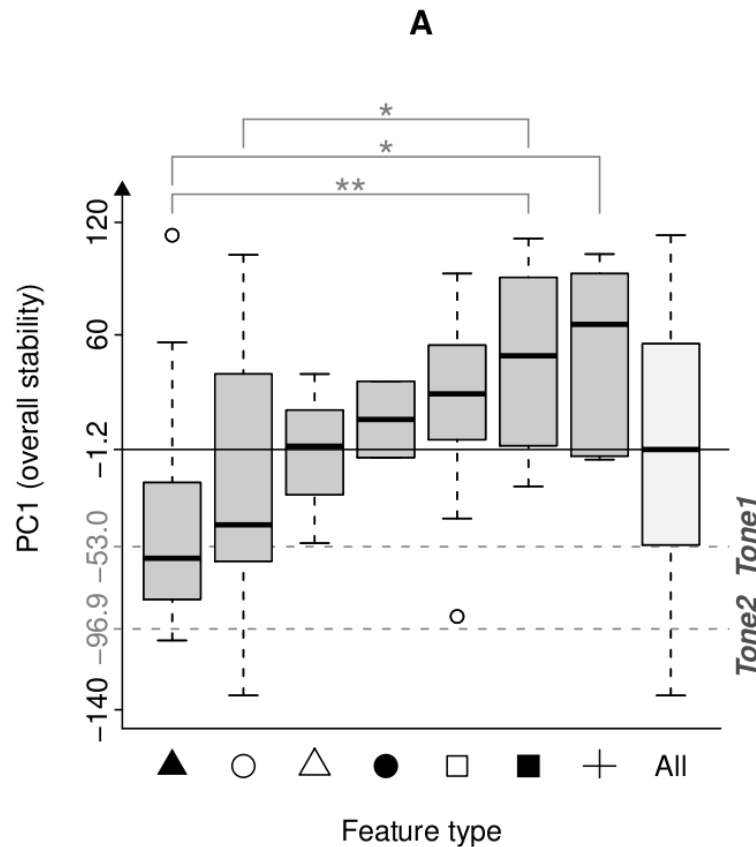
- **Similar results across:**
 - Lg. classifications: $0.96 \leq r \leq 0.99$, $p < 10^{-10}$
 - Outgroups: $0.49 \leq r \leq 0.92$, $p < 10^{-6}$, mean $r = 0.78$, PC_1 79%
- **Binary coding:**
 - Strong agreement: $0.59 \leq r \leq 0.98$, $p < 10^{-8}$, mean $r = 0.78$, PC_1 81.4%
 - *Tone2* (**8** of 86), *Tone1* (**23** of 86)
- **Polymorphic coding:**
 - Strong agreement: $0.51 \leq r \leq 0.99$, $p < 10^{-5}$, mean $r = 0.71$, PC_1 76.1%
 - *Tone* (**8** of 68)



Phylogenetic stability

Binary vs polymorphic codings

mean $r=0.61$, median $p=6.5 \cdot 10^{-9}$, PC_1 (agreement) 67.4%, PC_2 (bin vs poly) 16.1%



▲ = Word Order, ○ = Phonology, △ = Nominal Syntax, ● = Morphology, □ = Verbal Categories, ■ = Nominal Categories, + = Simple Clauses, and All = all types of features combined.



Phylogenetic stability

Results & conclusion

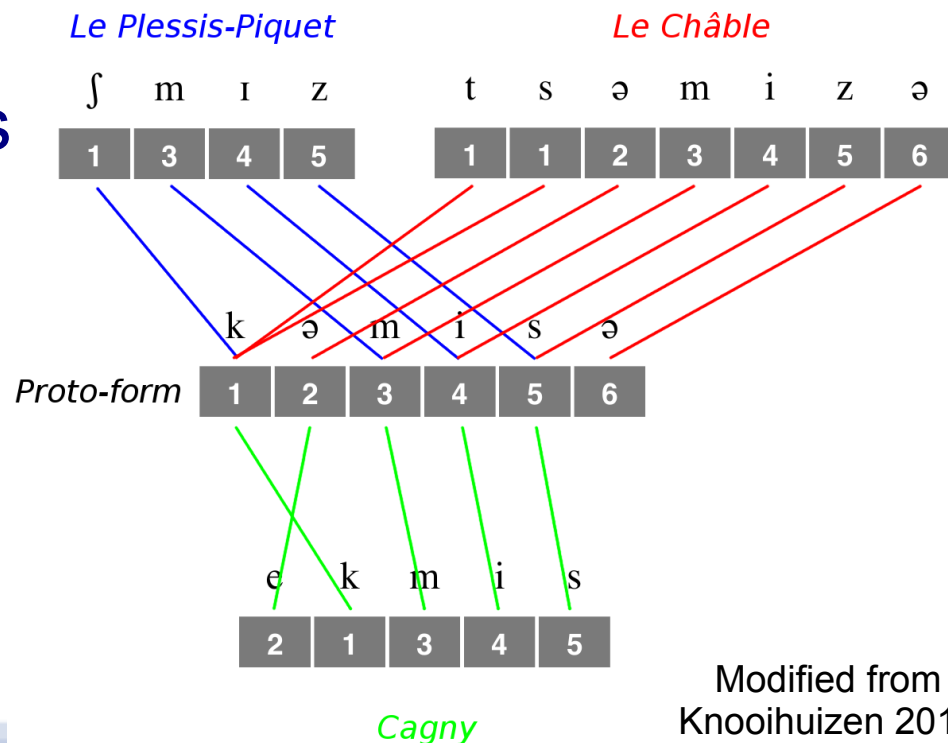
- *Tone* is **very stable** as **poly feat** ($t_{56}=9.7$, $p=1.35\cdot 10^{-13}$)
- *Tone2* is **very stable** as **bin feat** ($t_{70}=12.04$, $p<2.2\cdot 10^{-16}$) and **overall** ($t_{70}=12.27$, $p<2.2\cdot 10^{-16}$)
- *Tone1* **relatively stable** as **bin feat** ($t_{70}=4.35$, $p=4.5\cdot 10^{-5}$) and **overall** ($t_{70}=4.35$, $p=6.7\cdot 10^{-9}$)
- Many **caveats** (data quality and quantity, appropriateness of strict phylogenetic methods to typology, language isolates as outgroups, etc.) → **suggestive at best!!!**



Dialect variation

The dynamics of English dialects

- W. Maguire, A. McMahon & P. Heggarty; **methodology**
- 110 words; 36 varieties; “typical”, “traditional” & “emergent”
- narrow phonetic transcription → Heggarty's **slot-matching algorithm**:
 - *distance matrix* words & varieties
- *Networks* (NeighborNet)
- *MDS, word-level analysis*
- *Randomization tests*



Modified from
Knooihuizen 2010



Dialect variation

Language shift in Early Modern Dunkirk

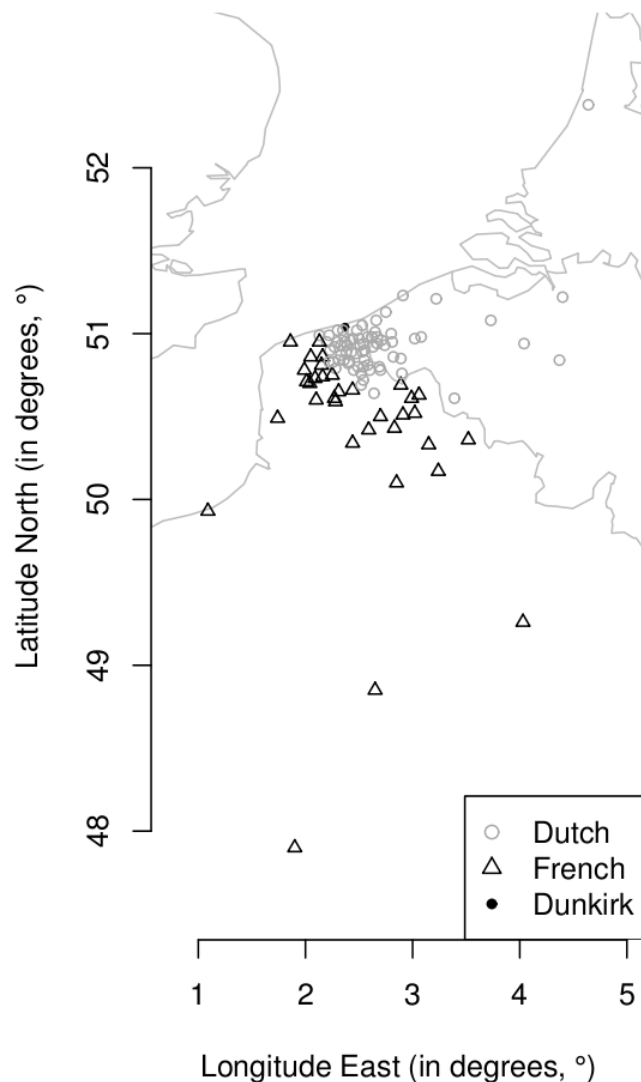
- R. Knooihuizen – also **methodology** only
- 2 related studies: language history + present dialects
- (a) **Demography and language shift:**
- Dunkirk: annexed 1662
- Marriage registers 1647, 1657, 1667, 1677, 1687 & 1697
- Parish of origin, name, priest's name, literacy, “age”
- → migration and marriage patterns



Dialect variation

EM Dunkirk

Locations by language



- *Point-pattern analysis* of parishes of origin vs language, gender and year
- Random Labeling Hypothesis (RLH)
- RLH rejected for language :) but not for gender nor year
- Rates of immigration by gender, year and language, ...
- Bias against Mixed marriages



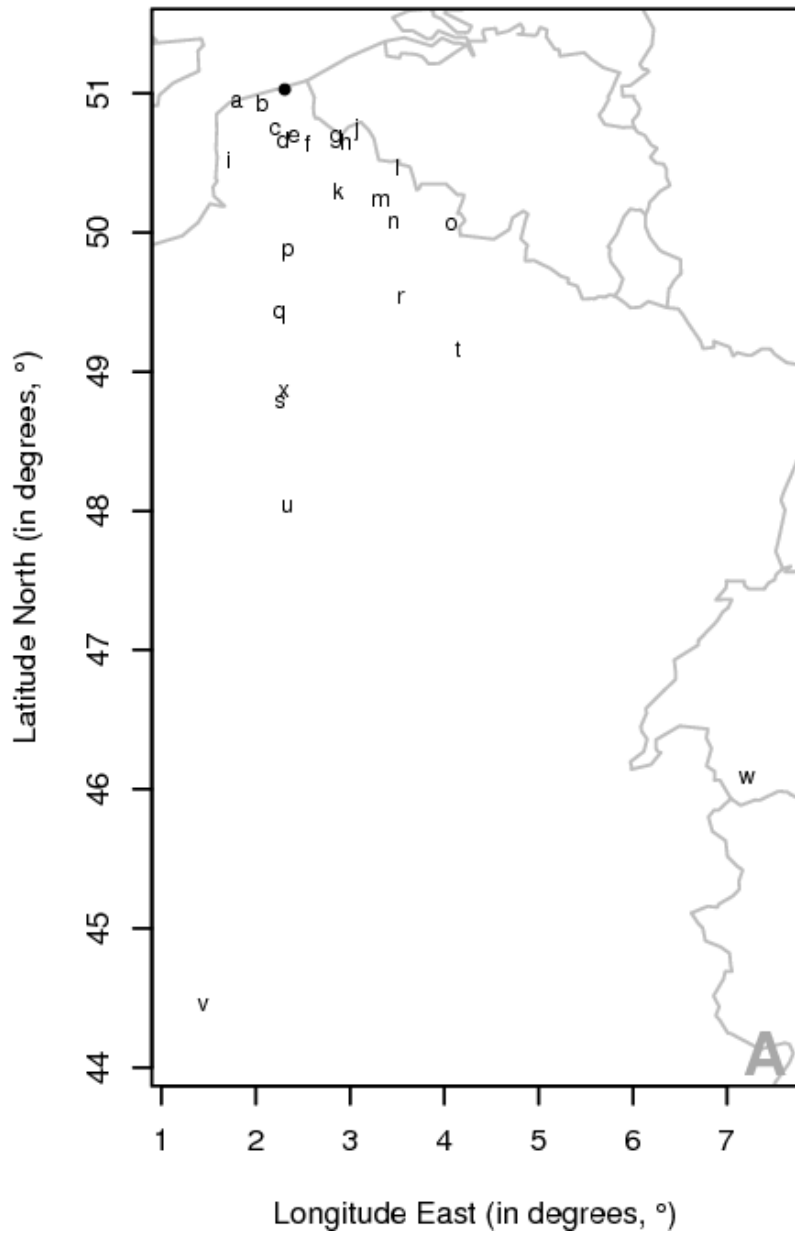
Dialect variation

Dunkirk: Present-day dialects

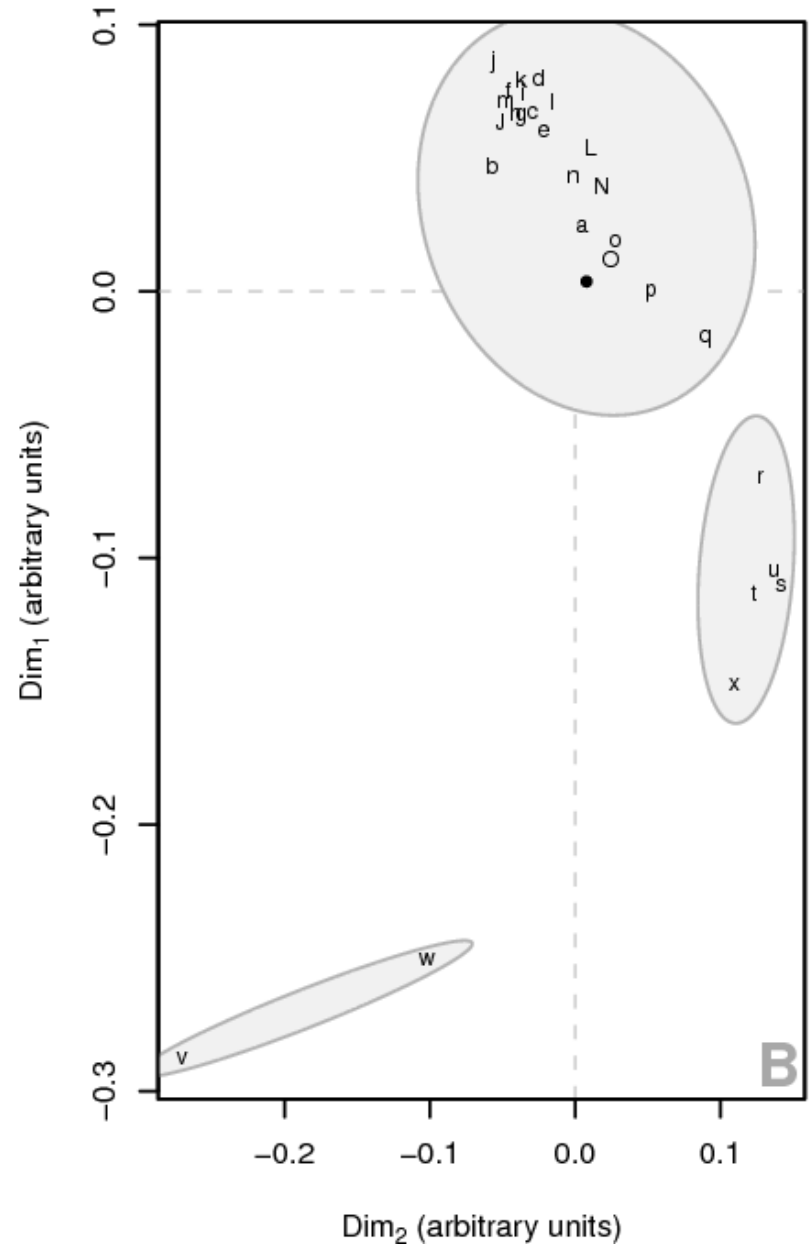
- (b) **Dunkirk dialect vs. other French varieties**
- Language atlases, 60 words, Heggarty's slot-matching algorithm
- *Network analysis*
- *Mantel correlation with geography*
- *MDS and hierarchical clustering*
- → Dunkirk is clearly a **Picard variety** but with **more distant influences**, possibly signaling earlier migration patterns?



Geographic locations



Multidimensional scaling of linguistic distances



Conclusions

- There are complex relationships between languages and genes on multiple levels/scales
- There are various quantitative methods which can be adapted to language data
- But there are also many assumptions which must be checked (and changed) against properties of language
- Collaborations between language scientists and biologists are bound to be fruitful for both sides
- However, communication is often difficult due to many factors (training, historical accidents, mistrust, ...)



End

... for More Info:

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *PNAS*, **104**, 10944-10949.

Ladd, D. R., Dediu, D., & Kinsella, A. R. (2008). Languages and genes: reflections on biolinguistics and the nature-nurture question. *Biolinguistics*, **2**(1), 114-126.

Dediu, D. (2008). The role of genetic biases in shaping the correlations between languages and genes. *Journal of Theoretical Biology*, **254**, 400-407.

Dediu, D. (2009). Genetic biasing through cultural transmission: Do simple Bayesian models of language evolution generalize? *Journal of Theoretical Biology*, **259**, 552-561.

Dediu, D. (2010). A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proc. Royal Society B*. Advance online publication. doi:10.1098/rspb.2010.1595.

Maguire, W., McMahon, A., Heggarty, P., & Dediu, D. (2010). The past, present, and future of English dialects: Quantifying convergence, divergence, and dynamic equilibrium. *Language Variation and Change*, **22**, 69-104.

Knooihuizen, R. & Dediu, D (*in preparation*). demographic factors in historical sociolinguistics: immigration, integration and language shift in early modern Dunkirk.

... Many Thanks to:

Bob Ladd, Morten Christiansen, Daniel Nettle, Kenny Smith, Simon Kirby, Jim Hurford, Fiona Jordan, Gwen Hyslop, Steve Levinson, Michael Cysouw, Michael Dunn, Russel Gray, Simon Greenhill and Alexandra Dima.

The University of Edinburgh, UK; ESRC UK; The Max Planck Institute for Psycholinguistics, Nijmegen

And to the organizers of BEDLAN 2010

