

Friends

Manja Marz

University of Marburg

Bled, Germany
19.02.2011

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

$$1025 \cdot 110 = 112,750 \text{ files}$$

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files
31,728,761 blasthits

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files

31,728,761 blasthits

Total 2,29 GB

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files

31,728,761 blasthits

Total 2,29 GB ~ 500,000 A4-pages

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files

31,728,761 blasthits

Total 2,29 GB ~ 500,000 A4-pages (500 books – > one library)

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files

31,728,761 blasthits

Total 2,29 GB ~ 500,000 A4-pages (500 books – > one library)
2 weeks

Science in Marburg

- Infernal Rfam v.10.0 on pig: 19 cpu years
- Blast Rfam on 110 metazoans (Evalue $< 10^{-4}$)

1025 · 110 = 112,750 files

31,728,761 blasthits

Total 2,29 GB ~ 500,000 A4-pages (500 books – > one library)
2 weeks and many weeks free time ;)

Acknowledgements

Thx 2:

Steffi, Marcus

C. germanicus

Thank You!

Introduction into the wide, complicated field of Molecular Biology:

Introduction into the wide, complicated field of Molecular Biology:

Feature Observation





































Biologists vs. Computer Scientists

Biologists: Favourite gene ... or hype scenario

Biologists: Favourite gene ... or hype scenario



Examples from Biology: Gene e and gene f

Examples from Biology: Gene *e* and gene *f*

Gene *e*

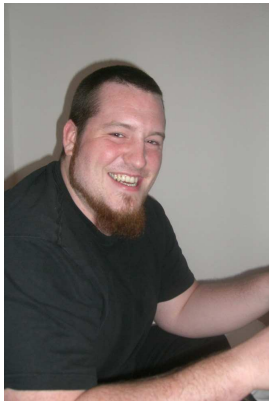


Examples from Biology: Gene *e* and gene *f*

Gene *e*



Gene *f*



Starting accurate analysis: $\Delta(e, f) = \Delta(f, e)$

Starting accurate analysis: $\Delta(e, f) = \Delta(f, e)$







$\forall x \in G : \Delta(e, f) < \Delta(e, x) \text{ and } \Delta(e, f) < \Delta(f, x)$

$\forall x \in G : \Delta(e, f) < \Delta(e, x)$ and $\Delta(e, f) < \Delta(f, x)$



$\forall x \in G : \Delta(e, f) < \Delta(e, x)$ and $\Delta(e, f) < \Delta(f, x)$



Sense, swapped



Sense, swapped



Sense, "Minus strand"



Antisense



Clustered with other genes



Aims

- Find conserved gene linkages throughout metazoans
- Determine time of linkage investigation
- Duplications
- Constant distances and orientations?
- Identification of divergent genes

Problem: A gene $s \in G$ jumps in: $\Delta(e, s) < \Delta(e, f)$ and $\Delta(f, s) < \Delta(e, f)$

Problem: A gene $s \in G$ jumps in: $\Delta(e, s) < \Delta(e, f)$ and $\Delta(f, s) < \Delta(e, f)$



Solution: Compare data in different time points



Solution: Compare data in different time points



Problem: another gene j overlaps: $\Delta(f, j) = 0$

Problem: another gene j overlaps: $\Delta(f, j) = 0$



Problem: another gene j overlaps: $\Delta(f, j) = 0$

Problem: another gene j overlaps: $\Delta(f, j) = 0$



Problem: Modified Gene or Pseudogene?

Problem: Modified Gene or Pseudogene?



Answered by evolution

Answered by evolution



Problem: Highly divergent genes: Fragment hit only

Problem: Highly divergent genes: Fragment hit only



Solved by the method itself



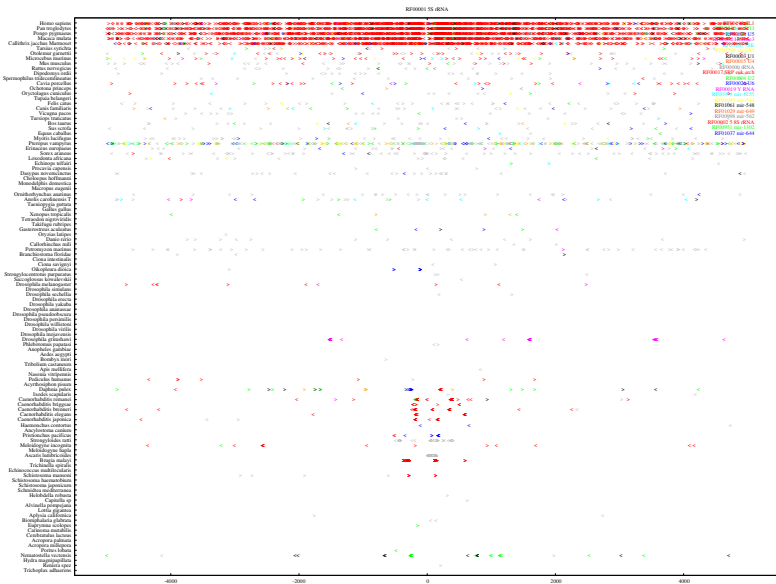
Problem: $\Delta(e, l) \ll \Delta(e, f)$

Problem: $\Delta(e, l) \ll \Delta(e, f)$

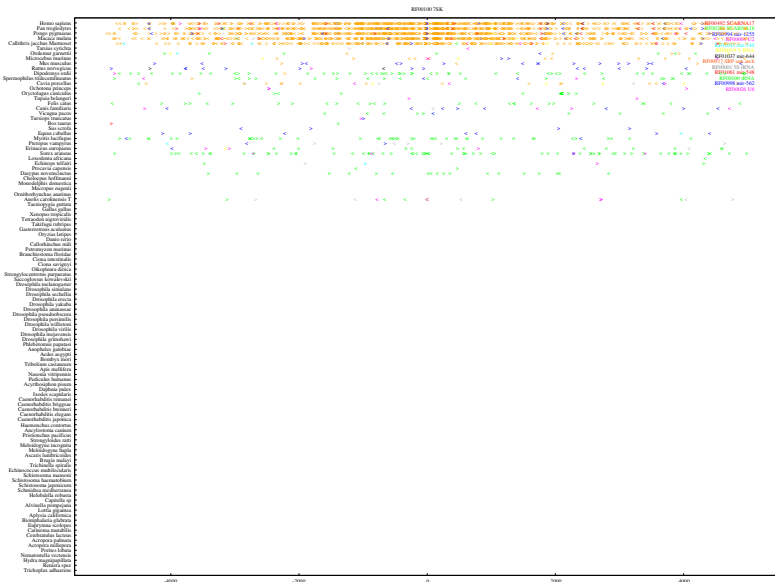


Some Preliminary Results: SCARNA-17-18

Some Preliminary Results: 5S RNA



Some Preliminary Results: 7SK RNA



Identification of a mobile element (gene p)

Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Identification of a mobile element (gene p)



Likelihood of a biological feature

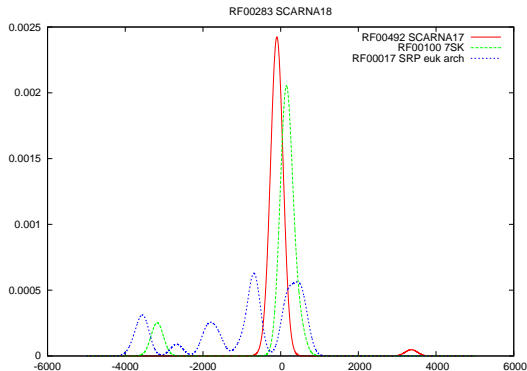
Assumptions:

- individual candidates for linkage between two genes are independently distributed
- each candidate can be modeled by a Gaussian $N(\mu, \sigma^2)$ with μ the distance between the two gene candidates and σ^2 chosen to be constant (independent of μ)

$$\mathcal{P}(I \geq k|n) = \sum_l \binom{n}{l} (4\sigma p)^k (1 - 4\sigma p)^{n-k}$$

confidence intervall $\phi = 0.9$

Statistically validation of SCARNA-17-18 linkage



RF00492 SCARNA17: $C_i = (-389,175)$, $iL = 564$

RF00100 7SK: $C_i = (-2883,756)$, $iL = 3639$

RF00017 SRP: $C_i = (-3613,679)$, $iL = 4292$

Conclusion

Conclusion

easy

Conclusion

easy
gain many data

Conclusion

easy
gain many data
complete analysis never done before

Conclusion

easy
gain many data
complete analysis never done before
biologists are impressed

Conclusion

easy
gain many data
complete analysis never done before
biologists are impressed

Conclusion

- Linkages of all known ncRNAs throughout all metazoans
- Correlation of Orientation and Distance
- Time point ncRNA linkage invention
- Occurrence in specific taxa?
- Identification of mobile elements
- Usage for divergent genes, pseudogenes
- Specific organism linkages and duplications

Acknowledgements

Thx 2:

e, f, j, l, p, s

Acknowledgements

Thx 2:

e, f, j, l, p, s

x