

BIOS 477/877 Bioinformatics and Molecular Evolution

Instructor: Etsuko Moriyama
(School of Biological Sciences)

Spring 2026

Lecture 5

BIOS477/877 L5 - 1

1

Today's topics

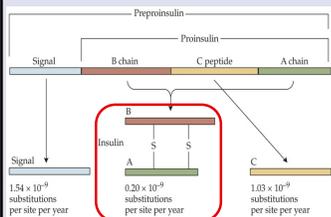
➤ Molecular Evolution - part 3

BIOS477/877 L5 - 2

2

How to identify selection

- Number of **synonymous** substitutions per site: d_S (or K_S)
- Number of **nonsynonymous** substitutions per site: d_N (or K_N)



$$d_N(AB) < d_N(C)$$

$$d_N(AB) = u_T(AB) * f_0(AB)$$

$$d_N(C) = u_T(C) * f_0(C)$$

Both mutation rates (u_T) and selective constraints (f_0) affect nonsynonymous rates (d_N)

Note: In literatures, K_a and K_s are also used instead of d_N and d_S

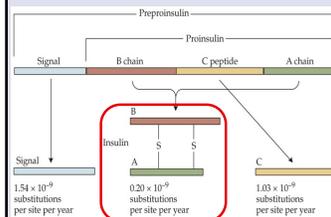
Based on nonsynonymous substitution rates estimated from human/rat comparison (divergence time: 80 million years ago) from Graur (2016)

BIOS477/877 L5 - 3

3

How to isolate u_T vs. f_0

- Number of **synonymous** substitutions per site: d_S (or K_S)
- Number of **nonsynonymous** substitutions per site: d_N (or K_N)



$$d_N(AB) < d_N(C)$$

$$u_T(AB) * f_0(AB) < u_T(C) * f_0(C)$$

What causes $d_N(AB) < d_N(C)$?

Both u_T and f_0 affect d_N :

$u_T(AB) < u_T(C)$?

$f_0(AB) < f_0(C)$?

or both?

Based on nonsynonymous substitution rates estimated from human/rat comparison (divergence time: 80 million years ago) from Graur (2016)

BIOS477/877 L5 - 4

4

How to isolate u_T vs. f_0 (continued)

$$d_N(AB) < d_N(C)$$

$$u_T(AB) * f_0(AB) < u_T(C) * f_0(C)$$

If mutation rates are constant within a gene,

$$\rightarrow u_T(AB) = u_T(C)$$

$\rightarrow d_N(AB) < d_N(C)$ can be explained by $f_0(AB) < f_0(C)$

If mutation rates are different,

$$\rightarrow u_T(AB) \neq u_T(C)$$

$\rightarrow d_N(AB) < d_N(C)$ cannot be simply explained by $f_0(AB) < f_0(C)$

BIOS477/877 L5 - 5

5

How to isolate f_0

Synonymous substitutions are assumed to be neutral or near-neutral:

$$f_0 \approx 1 \text{ (neutral)}$$

$$\rightarrow d_S = u_T f_0 \approx u_T$$

$$d_S(AB) = u_T(AB)$$

$$d_S(C) = u_T(C)$$

$$d_N(AB) = u_T(AB) * f_0(AB)$$

$$d_S(AB) = u_T(AB)$$

$$\frac{d_N(AB)}{d_S(AB)} = \frac{u_T(AB) * f_0(AB)}{u_T(AB)}$$

$$\frac{d_N(AB)}{d_S(AB)} = f_0(AB)$$

$$d_N(C) = u_T(C) * f_0(C)$$

$$d_S(C) = u_T(C)$$

$$\frac{d_N(C)}{d_S(C)} = \frac{u_T(C) * f_0(C)}{u_T(C)}$$

$$\frac{d_N(C)}{d_S(C)} = f_0(C)$$

BIOS477/877 L5 - 6

6

How to isolate f_0 (continued)

If synonymous substitutions are assumed to be neutral or near-neutral: $\frac{d_N(AB)}{d_S(AB)} = f_0(AB)$ and $\frac{d_N(C)}{d_S(C)} = f_0(C)$

d_N/d_S can be used to identify selection

- Shows only the level of selective constraints (f_0)
- We don't need to worry about mutation rates (u_T)
→ even if $u_T(AB) \neq u_T(C)$, d_N/d_S can be compared

BIOS477/877 L5 - 7

7

Identifying selection

d_N/d_S can be used to identify selection

$d_N(AB) < d_N(C)$

The better comparison is:
 $d_N(AB)/d_S(AB)$
vs $d_N(C)/d_S(C)$

Even if mutation rates are different, we can compare:
 $f_0(AB)$ vs. $f_0(C)$

Based on nonsynonymous substitution rates estimated from human/rat comparison (divergence time: 80 million years ago) from Graur (2016)

BIOS477/877 L5 - 8

8

When divergence time is different

$d_{N1} < d_{N2}, d_{S1} < d_{S2}$

This may be caused simply by the difference in divergence time ($t_1 < t_2$)

d_{N1}/d_{S1} vs. d_{N2}/d_{S2}

By comparing d_N/d_S , the time effect can be cancelled out.
We can compare the selective constraints between genes and between lineages.

$d_{N1} = u_{T1} * f_{01} * 2t_1, d_{S1} = u_{T1} * 2t_1 \rightarrow d_{N1}/d_{S1} = u_{T1} * f_{01} * 2t_1 / (u_{T1} * 2t_1)$
 $d_{N2} = u_{T2} * f_{02} * 2t_2, d_{S2} = u_{T2} * 2t_2 \rightarrow d_{N2}/d_{S2} = u_{T2} * f_{02} * 2t_2 / (u_{T2} * 2t_2)$

BIOS477/877 L5 - 9

9

Identifying selection

If nonsynonymous (replacement) substitutions are neutral:

$d_N > d_S$ or $d_N \approx d_S$ or $d_N < d_S$?

BIOS477/877 L5 - 10

10

Identifying selection (continued)

If nonsynonymous (replacement) substitutions are neutral:

$d_N = u_T * f_0$, where $f_0 \approx 1$

→ $d_N = u_T * f_0 \approx u_T$

Since $d_S = u_T \rightarrow d_N \approx d_S$

BIOS477/877 L5 - 11

11

Identifying selection (continued)

If nonsynonymous (replacement) substitutions are neutral:

$d_N \approx d_S \rightarrow d_N/d_S \approx 1$

d_S is used as a control (substitution rate at neutral or near neutral)

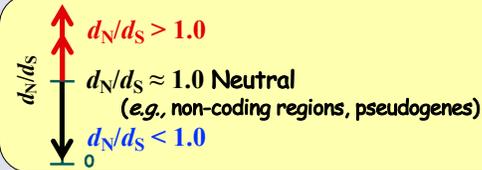
→ d_N is evaluated against d_S

BIOS477/877 L5 - 12

12

Identifying selection (continued)

d_N/d_S can be used to identify selection



BIOS477/877 L5 - 13

13

Identifying selection (continued)

d_N/d_S can be used to identify selection

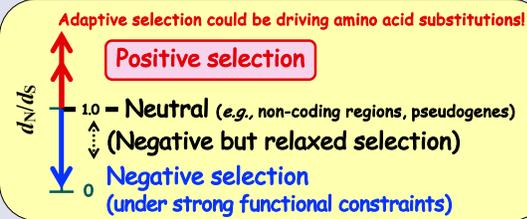


BIOS477/877 L5 - 14

14

Type of selection (continued)

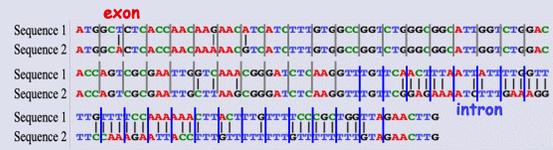
d_N/d_S can be used to identify selection



BIOS477/877 L5 - 16

16

Identifying selection (exon vs. intron)



| | Exon | | | Intron | | |
|-------------------------------|------|-----|------|--------|------|------|
| [Codon position] | 1st | 2nd | 3rd | 1st | 2nd | 3rd |
| # nucleotide sites | 31 | 31 | 31 | 23 | 22 | 22 |
| # nucleotide substitutions | 2 | 0 | 4 | 12 | 12 | 14 |
| Nucleotide substitutions/site | 0.06 | 0 | 0.13 | 0.52 | 0.55 | 0.64 |

Exon(2nd) < Exon(3rd) Intron(2nd) ≈ or ≈ Intron(3rd)

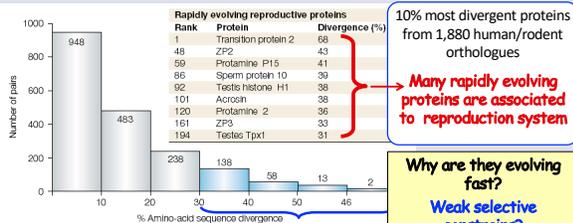
Exon(2nd) / Exon(3rd) = 0 Intron(2nd) / Intron(3rd) = 0.86 (≈1)

(Similar to d_N/d_S analysis)

BIOS477/877 L4 - 17

17

Positive selection (reproductive proteins)



10% most divergent proteins from 1,880 human/rodent orthologues
 Many rapidly evolving proteins are associated to reproduction system

Why are they evolving fast?
 Weak selective constrains? (relaxed selection)
 or
 Divergence is favored? (positive selection)

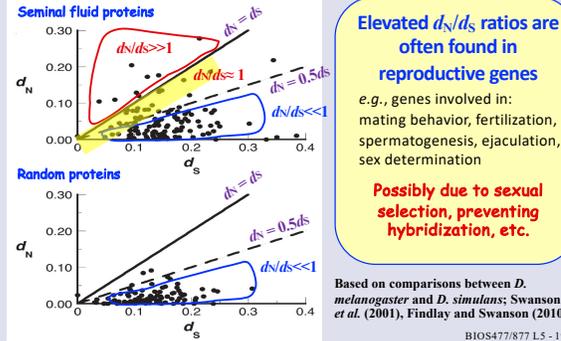
Figure 1 | Rapidly evolving proteins. Comparison of 1,880 human-rodent orthologues from Makalowski & Boguski² plotted as a frequency of the occurrence of genes with a varying percentage of amino acid divergence. The portion that contains the 10% most divergent proteins is shown in blue; reproductive proteins that are among the 10% most divergent proteins are listed in blue. Testis-specific protein 1, ZP2/3, zona pellucida 2/3.

From Swanson & Vacquier (2002)

BIOS477/877 L5 - 18

18

Positive selection (reproductive proteins)



Elevated d_N/d_S ratios are often found in reproductive genes
 e.g., genes involved in: mating behavior, fertilization, spermatogenesis, ejaculation, sex determination

Possibly due to sexual selection, preventing hybridization, etc.

Based on comparisons between *D. melanogaster* and *D. simulans*; Swanson et al. (2001), Findlay and Swanson (2010)

BIOS477/877 L5 - 19

19

Proteins under positive selection

Table 1. Selected examples of protein-coding genes in which positive selection was detected by using the d_N/d_S ratio

| Gene | Species | Ref. |
|--|--|------|
| Defensive systems or immunity | | |
| Genes involved in defense | | |
| DNAI1 (retroviral gene) | Escherichia coli | 61 |
| Dufin genes | Rodents | 62 |
| Defensin genes | Escherichia coli | 63 |
| Fv1 | Mus | 64 |
| Immunoglobulin V _H genes | Mammals | 65 |
| MHC genes | Mammals | 66 |
| Polynucleotidase inhibitor genes | Primates and rodents | 67 |
| Rh blood group and RH50 genes | Primates and rodents | 68 |
| Rhesus-like genes | Primates | 69 |
| Transferrin gene | Salmonid fishes | 70 |
| Type I interferon gene | Mammals | 71 |
| α -Proteinase inhibitor genes | Rodents | 72 |
| Reproduction | | |
| Capid gene | FMD virus | 73 |
| CSP, TRAP, MSA2 and PRS3 | Flammarion (fish) | 74 |
| Dalmanite coding region | Hepatitis D virus | 75 |
| E gene | Phage G4 (bacteriophage) | 76 |
| Digestion | | |
| ATP synthase F ₁ subunit gene | Escherichia coli | 77 |
| Toxin proteins | | |
| Envelopin | ATP synthase F ₁ subunit gene | 78 |
| gH glycoprotein | Herpesvirus | 79 |
| HerpesgD | Herpesvirus | 80 |
| Invasin | Herpesvirus | 81 |
| Membran | Herpesvirus | 82 |
| msp 1a | Angiostroma marginale | 83 |
| msp 1b | Herpesvirus | 84 |
| msp 1c | Herpesvirus | 85 |
| msp 1d | Herpesvirus | 86 |
| msp 1e | Herpesvirus | 87 |
| msp 1f | Herpesvirus | 88 |
| msp 1g | Herpesvirus | 89 |
| msp 1h | Herpesvirus | 90 |
| msp 1i | Herpesvirus | 91 |
| msp 1j | Herpesvirus | 92 |
| msp 1k | Herpesvirus | 93 |
| msp 1l | Herpesvirus | 94 |
| msp 1m | Herpesvirus | 95 |
| msp 1n | Herpesvirus | 96 |
| msp 1o | Herpesvirus | 97 |
| msp 1p | Herpesvirus | 98 |
| msp 1q | Herpesvirus | 99 |
| msp 1r | Herpesvirus | 100 |
| msp 1s | Herpesvirus | 101 |
| msp 1t | Herpesvirus | 102 |
| msp 1u | Herpesvirus | 103 |
| msp 1v | Herpesvirus | 104 |
| msp 1w | Herpesvirus | 105 |
| msp 1x | Herpesvirus | 106 |
| msp 1y | Herpesvirus | 107 |
| msp 1z | Herpesvirus | 108 |
| msp 2 | Herpesvirus | 109 |
| msp 3 | Herpesvirus | 110 |
| msp 4 | Herpesvirus | 111 |
| msp 5 | Herpesvirus | 112 |
| msp 6 | Herpesvirus | 113 |
| msp 7 | Herpesvirus | 114 |
| msp 8 | Herpesvirus | 115 |
| msp 9 | Herpesvirus | 116 |
| msp 10 | Herpesvirus | 117 |
| msp 11 | Herpesvirus | 118 |
| msp 12 | Herpesvirus | 119 |
| msp 13 | Herpesvirus | 120 |
| msp 14 | Herpesvirus | 121 |
| msp 15 | Herpesvirus | 122 |
| msp 16 | Herpesvirus | 123 |
| msp 17 | Herpesvirus | 124 |
| msp 18 | Herpesvirus | 125 |
| msp 19 | Herpesvirus | 126 |
| msp 20 | Herpesvirus | 127 |
| msp 21 | Herpesvirus | 128 |
| msp 22 | Herpesvirus | 129 |
| msp 23 | Herpesvirus | 130 |
| msp 24 | Herpesvirus | 131 |
| msp 25 | Herpesvirus | 132 |
| msp 26 | Herpesvirus | 133 |
| msp 27 | Herpesvirus | 134 |
| msp 28 | Herpesvirus | 135 |
| msp 29 | Herpesvirus | 136 |
| msp 30 | Herpesvirus | 137 |
| msp 31 | Herpesvirus | 138 |
| msp 32 | Herpesvirus | 139 |
| msp 33 | Herpesvirus | 140 |
| msp 34 | Herpesvirus | 141 |
| msp 35 | Herpesvirus | 142 |
| msp 36 | Herpesvirus | 143 |
| msp 37 | Herpesvirus | 144 |
| msp 38 | Herpesvirus | 145 |
| msp 39 | Herpesvirus | 146 |
| msp 40 | Herpesvirus | 147 |
| msp 41 | Herpesvirus | 148 |
| msp 42 | Herpesvirus | 149 |
| msp 43 | Herpesvirus | 150 |
| msp 44 | Herpesvirus | 151 |
| msp 45 | Herpesvirus | 152 |
| msp 46 | Herpesvirus | 153 |
| msp 47 | Herpesvirus | 154 |
| msp 48 | Herpesvirus | 155 |
| msp 49 | Herpesvirus | 156 |
| msp 50 | Herpesvirus | 157 |
| msp 51 | Herpesvirus | 158 |
| msp 52 | Herpesvirus | 159 |
| msp 53 | Herpesvirus | 160 |
| msp 54 | Herpesvirus | 161 |
| msp 55 | Herpesvirus | 162 |
| msp 56 | Herpesvirus | 163 |
| msp 57 | Herpesvirus | 164 |
| msp 58 | Herpesvirus | 165 |
| msp 59 | Herpesvirus | 166 |
| msp 60 | Herpesvirus | 167 |
| msp 61 | Herpesvirus | 168 |
| msp 62 | Herpesvirus | 169 |
| msp 63 | Herpesvirus | 170 |
| msp 64 | Herpesvirus | 171 |
| msp 65 | Herpesvirus | 172 |
| msp 66 | Herpesvirus | 173 |
| msp 67 | Herpesvirus | 174 |
| msp 68 | Herpesvirus | 175 |
| msp 69 | Herpesvirus | 176 |
| msp 70 | Herpesvirus | 177 |
| msp 71 | Herpesvirus | 178 |
| msp 72 | Herpesvirus | 179 |
| msp 73 | Herpesvirus | 180 |
| msp 74 | Herpesvirus | 181 |
| msp 75 | Herpesvirus | 182 |
| msp 76 | Herpesvirus | 183 |
| msp 77 | Herpesvirus | 184 |
| msp 78 | Herpesvirus | 185 |
| msp 79 | Herpesvirus | 186 |
| msp 80 | Herpesvirus | 187 |
| msp 81 | Herpesvirus | 188 |
| msp 82 | Herpesvirus | 189 |
| msp 83 | Herpesvirus | 190 |
| msp 84 | Herpesvirus | 191 |
| msp 85 | Herpesvirus | 192 |
| msp 86 | Herpesvirus | 193 |
| msp 87 | Herpesvirus | 194 |
| msp 88 | Herpesvirus | 195 |
| msp 89 | Herpesvirus | 196 |
| msp 90 | Herpesvirus | 197 |
| msp 91 | Herpesvirus | 198 |
| msp 92 | Herpesvirus | 199 |
| msp 93 | Herpesvirus | 200 |
| msp 94 | Herpesvirus | 201 |
| msp 95 | Herpesvirus | 202 |
| msp 96 | Herpesvirus | 203 |
| msp 97 | Herpesvirus | 204 |
| msp 98 | Herpesvirus | 205 |
| msp 99 | Herpesvirus | 206 |
| msp 100 | Herpesvirus | 207 |
| msp 101 | Herpesvirus | 208 |
| msp 102 | Herpesvirus | 209 |
| msp 103 | Herpesvirus | 210 |
| msp 104 | Herpesvirus | 211 |
| msp 105 | Herpesvirus | 212 |
| msp 106 | Herpesvirus | 213 |
| msp 107 | Herpesvirus | 214 |
| msp 108 | Herpesvirus | 215 |
| msp 109 | Herpesvirus | 216 |
| msp 110 | Herpesvirus | 217 |
| msp 111 | Herpesvirus | 218 |
| msp 112 | Herpesvirus | 219 |
| msp 113 | Herpesvirus | 220 |
| msp 114 | Herpesvirus | 221 |
| msp 115 | Herpesvirus | 222 |
| msp 116 | Herpesvirus | 223 |
| msp 117 | Herpesvirus | 224 |
| msp 118 | Herpesvirus | 225 |
| msp 119 | Herpesvirus | 226 |
| msp 120 | Herpesvirus | 227 |
| msp 121 | Herpesvirus | 228 |
| msp 122 | Herpesvirus | 229 |
| msp 123 | Herpesvirus | 230 |
| msp 124 | Herpesvirus | 231 |
| msp 125 | Herpesvirus | 232 |
| msp 126 | Herpesvirus | 233 |
| msp 127 | Herpesvirus | 234 |
| msp 128 | Herpesvirus | 235 |
| msp 129 | Herpesvirus | 236 |
| msp 130 | Herpesvirus | 237 |
| msp 131 | Herpesvirus | 238 |
| msp 132 | Herpesvirus | 239 |
| msp 133 | Herpesvirus | 240 |
| msp 134 | Herpesvirus | 241 |
| msp 135 | Herpesvirus | 242 |
| msp 136 | Herpesvirus | 243 |
| msp 137 | Herpesvirus | 244 |
| msp 138 | Herpesvirus | 245 |
| msp 139 | Herpesvirus | 246 |
| msp 140 | Herpesvirus | 247 |
| msp 141 | Herpesvirus | 248 |
| msp 142 | Herpesvirus | 249 |
| msp 143 | Herpesvirus | 250 |
| msp 144 | Herpesvirus | 251 |
| msp 145 | Herpesvirus | 252 |
| msp 146 | Herpesvirus | 253 |
| msp 147 | Herpesvirus | 254 |
| msp 148 | Herpesvirus | 255 |
| msp 149 | Herpesvirus | 256 |
| msp 150 | Herpesvirus | 257 |
| msp 151 | Herpesvirus | 258 |
| msp 152 | Herpesvirus | 259 |
| msp 153 | Herpesvirus | 260 |
| msp 154 | Herpesvirus | 261 |
| msp 155 | Herpesvirus | 262 |
| msp 156 | Herpesvirus | 263 |
| msp 157 | Herpesvirus | 264 |
| msp 158 | Herpesvirus | 265 |
| msp 159 | Herpesvirus | 266 |
| msp 160 | Herpesvirus | 267 |
| msp 161 | Herpesvirus | 268 |
| msp 162 | Herpesvirus | 269 |
| msp 163 | Herpesvirus | 270 |
| msp 164 | Herpesvirus | 271 |
| msp 165 | Herpesvirus | 272 |
| msp 166 | Herpesvirus | 273 |
| msp 167 | Herpesvirus | 274 |
| msp 168 | Herpesvirus | 275 |
| msp 169 | Herpesvirus | 276 |
| msp 170 | Herpesvirus | 277 |
| msp 171 | Herpesvirus | 278 |
| msp 172 | Herpesvirus | 279 |
| msp 173 | Herpesvirus | 280 |
| msp 174 | Herpesvirus | 281 |
| msp 175 | Herpesvirus | 282 |
| msp 176 | Herpesvirus | 283 |
| msp 177 | Herpesvirus | 284 |
| msp 178 | Herpesvirus | 285 |
| msp 179 | Herpesvirus | 286 |
| msp 180 | Herpesvirus | 287 |
| msp 181 | Herpesvirus | 288 |
| msp 182 | Herpesvirus | 289 |
| msp 183 | Herpesvirus | 290 |
| msp 184 | Herpesvirus | 291 |
| msp 185 | Herpesvirus | 292 |
| msp 186 | Herpesvirus | 293 |
| msp 187 | Herpesvirus | 294 |
| msp 188 | Herpesvirus | 295 |
| msp 189 | Herpesvirus | 296 |
| msp 190 | Herpesvirus | 297 |
| msp 191 | Herpesvirus | 298 |
| msp 192 | Herpesvirus | 299 |
| msp 193 | Herpesvirus | 300 |
| msp 194 | Herpesvirus | 301 |
| msp 195 | Herpesvirus | 302 |
| msp 196 | Herpesvirus | 303 |
| msp 197 | Herpesvirus | 304 |
| msp 198 | Herpesvirus | 305 |
| msp 199 | Herpesvirus | 306 |
| msp 200 | Herpesvirus | 307 |
| msp 201 | Herpesvirus | 308 |
| msp 202 | Herpesvirus | 309 |
| msp 203 | Herpesvirus | 310 |
| msp 204 | Herpesvirus | 311 |
| msp 205 | Herpesvirus | 312 |
| msp 206 | Herpesvirus | 313 |
| msp 207 | Herpesvirus | 314 |
| msp 208 | Herpesvirus | 315 |
| msp 209 | Herpesvirus | 316 |
| msp 210 | Herpesvirus | 317 |
| msp 211 | Herpesvirus | 318 |
| msp 212 | Herpesvirus | 319 |
| msp 213 | Herpesvirus | 320 |
| msp 214 | Herpesvirus | 321 |
| msp 215 | Herpesvirus | 322 |
| msp 216 | Herpesvirus | 323 |
| msp 217 | Herpesvirus | 324 |
| msp 218 | Herpesvirus | 325 |
| msp 219 | Herpesvirus | 326 |
| msp 220 | Herpesvirus | 327 |
| msp 221 | Herpesvirus | 328 |
| msp 222 | Herpesvirus | 329 |
| msp 223 | Herpesvirus | 330 |
| msp 224 | Herpesvirus | 331 |
| msp 225 | Herpesvirus | 332 |
| msp 226 | Herpesvirus | 333 |
| msp 227 | Herpesvirus | 334 |
| msp 228 | Herpesvirus | 335 |
| msp 229 | Herpesvirus | 336 |
| msp 230 | Herpesvirus | 337 |
| msp 231 | Herpesvirus | 338 |
| msp 232 | Herpesvirus | 339 |
| msp 233 | Herpesvirus | 340 |
| msp 234 | Herpesvirus | 341 |
| msp 235 | Herpesvirus | 342 |
| msp 236 | Herpesvirus | 343 |
| msp 237 | Herpesvirus | 344 |
| msp 238 | Herpesvirus | 345 |
| msp 239 | Herpesvirus | 346 |
| msp 240 | Herpesvirus | 347 |
| msp 241 | Herpesvirus | 348 |
| msp 242 | Herpesvirus | 349 |
| msp 243 | Herpesvirus | 350 |
| msp 244 | Herpesvirus | 351 |
| msp 245 | Herpesvirus | 352 |
| msp 246 | Herpesvirus | 353 |
| msp 247 | Herpesvirus | 354 |
| msp 248 | Herpesvirus | 355 |
| msp 249 | Herpesvirus | 356 |
| msp 250 | Herpesvirus | 357 |
| msp 251 | Herpesvirus | 358 |
| msp 252 | Herpesvirus | 359 |
| msp 253 | Herpesvirus | 360 |
| msp 254 | Herpesvirus | 361 |
| msp 255 | Herpesvirus | 362 |
| msp 256 | Herpesvirus | 363 |
| msp 257 | Herpesvirus | 364 |
| msp 258 | Herpesvirus | 365 |
| msp 259 | Herpesvirus | 366 |
| msp 260 | Herpesvirus | 367 |
| msp 261 | Herpesvirus | 368 |
| msp 262 | Herpesvirus | 369 |
| msp 263 | Herpesvirus | 370 |
| msp 264 | Herpesvirus | 371 |
| msp 265 | Herpesvirus | 372 |
| msp 266 | Herpesvirus | 373 |
| msp 267 | Herpesvirus | 374 |
| msp 268 | Herpesvirus | 375 |
| msp 269 | Herpesvirus | 376 |
| msp 270 | Herpesvirus | 377 |
| msp 271 | Herpesvirus | 378 |
| msp 272 | Herpesvirus | 379 |
| msp 273 | Herpesvirus | 380 |
| msp 274 | Herpesvirus | 381 |
| msp 275 | Herpesvirus | 382 |
| msp 276 | Herpesvirus | 383 |
| msp 277 | Herpesvirus | 384 |
| msp 278 | Herpesvirus | 385 |
| msp 279 | Herpesvirus | 386 |
| msp 280 | Herpesvirus | 387 |
| msp 281 | Herpesvirus | 388 |
| msp 282 | Herpesvirus | 389 |
| msp 283 | Herpesvirus | 390 |
| msp 284 | Herpesvirus | 391 |
| msp 285 | Herpesvirus | 392 |
| msp 286 | | |

Detection of selection and data quality

The quality of the sequence, the degree of misannotation, and ambiguities in the multiple sequence alignment, all affect identification of positive selection!

Table 1
Inferred Percentage of PSGs as a Function of Sequencing Coverage (PSG: positively selected genes)

| | Higher quality → Coverage $\geq 3\times$ | | | Coverage $< 3\times$ ← Lower quality | | | $P(\chi^2)$ |
|---------|--|-----|-------|--------------------------------------|-------|-------|------------------------|
| | Total | PSG | % PSG | Total | PSG | % PSG | |
| Human | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Chimp | 1,144 | 9 | 0.8 | 1,836 | 74 | 4.0 | 9.6×10^{-8} |
| Macaque | 896 | 32 | 3.6 | 2,084 | 488 | 23.4 | 8.1×10^{-46} |
| Mouse | 2,493 | 77 | 3.1 | 487 | 37 | 7.6 | 1.3×10^{-6} |
| Rat | 1,841 | 93 | 5.1 | 1,139 | 217 | 19.1 | 3.2×10^{-38} |
| Dog | 1,568 | 212 | 13.5 | 1,412 | 481 | 34.1 | 1.7×10^{-49} |
| Cow | 1,086 | 54 | 5.0 | 1,894 | 249 | 13.1 | 4.8×10^{-14} |
| Total | 9,028 | 477 | 5.3 | 8,852 | 1,546 | 17.5 | 2.3×10^{-166} |

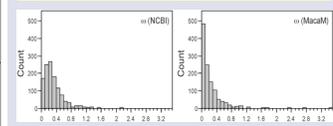
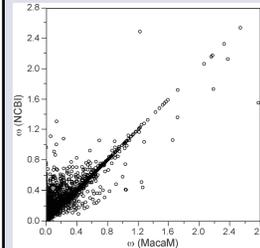
Schneider *et al.* (2009)

BIOS477/877 L5 - 26

26

Detection of selection and data quality (continued)

Low quality annotation of the rhesus macaque genome inflated ω (d_N/d_S)



• Comparisons: human vs. rhesus genes

• Genome annotations:

- NCBI: draft quality rhesus genome
- MacaM: new high quality rhesus genome (same assembly but different gene annotation)

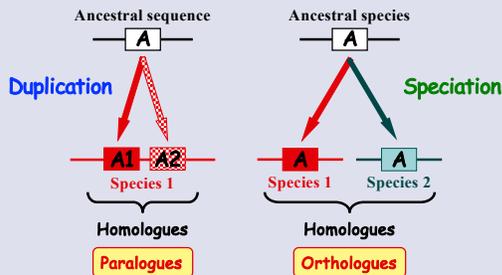
Gradnigo *et al.* (2016)

BIOS477/877 L5 - 27

27

Homologue, orthologue, paralogue

➤ **Homologues:** sequences that share a common ancestor



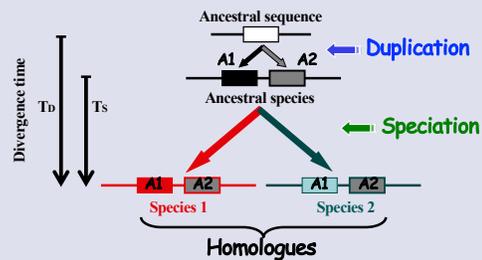
Read Theißen (2002) [very short review]
Gabaldon & Koonin (2013)

BIOS477/877 L5 - 28

28

Homologue, orthologue, paralogue (a more complex gene history)

➤ **Homologues:** sequences that share a common ancestor

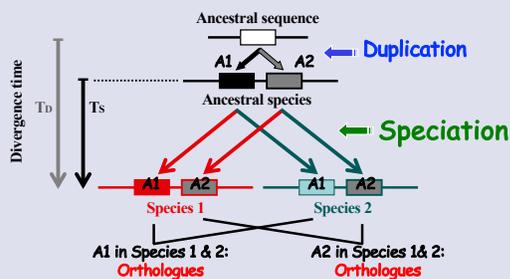


BIOS477/877 L5 - 29

29

Homologue, orthologue, paralogue (identifying orthologues)

➤ **Orthologues:** derived from a **speciation** event

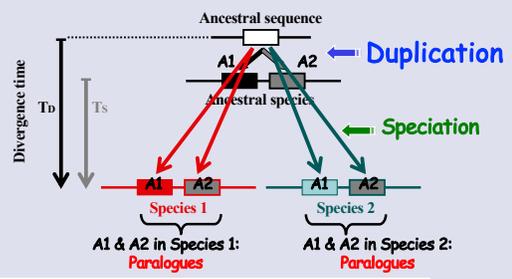


BIOS477/877 L5 - 30

30

Homologue, orthologue, paralogue (identifying paralogues)

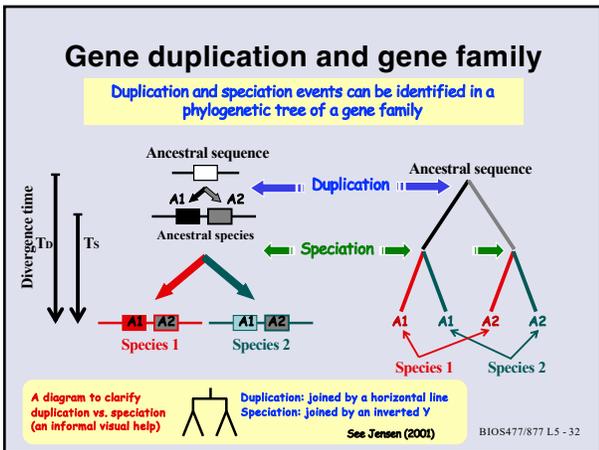
➤ **Paralogues:** derived from a **duplication** event



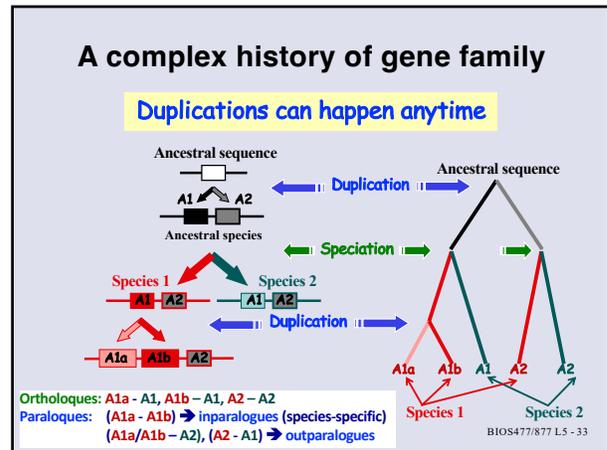
A1:Species 1 & A2:Species 2, A2:Species 1 & A1:Species 2 are also paralogues (paralogues can be in different genomes)

BIOS477/877 L5 - 31

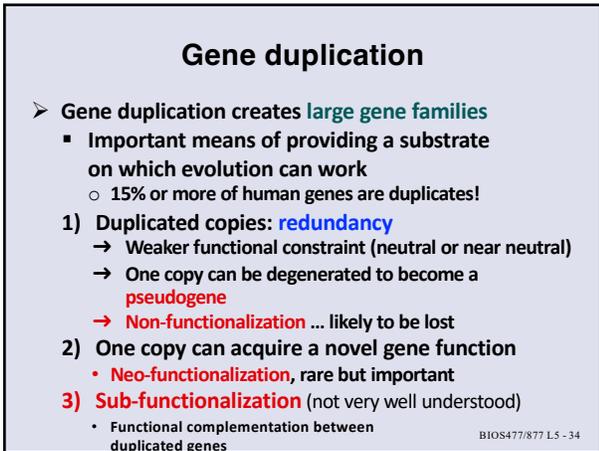
31



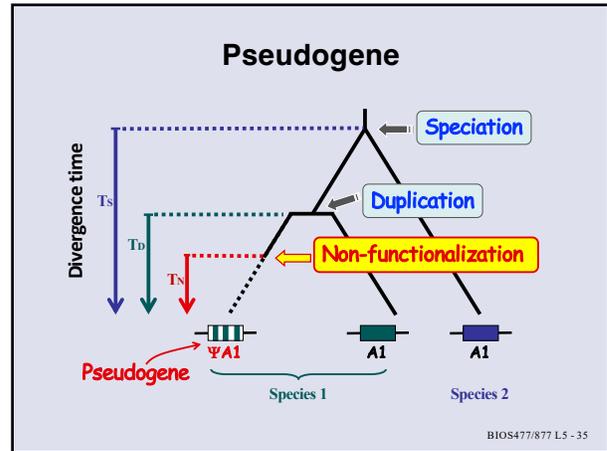
32



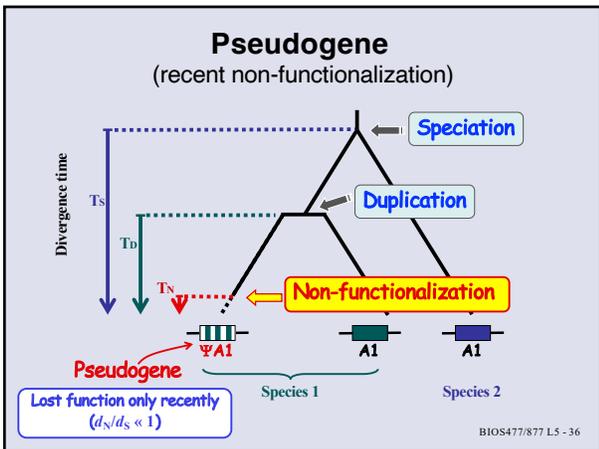
33



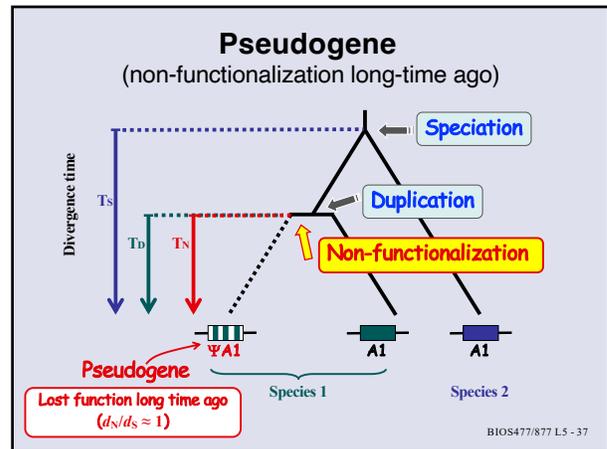
34



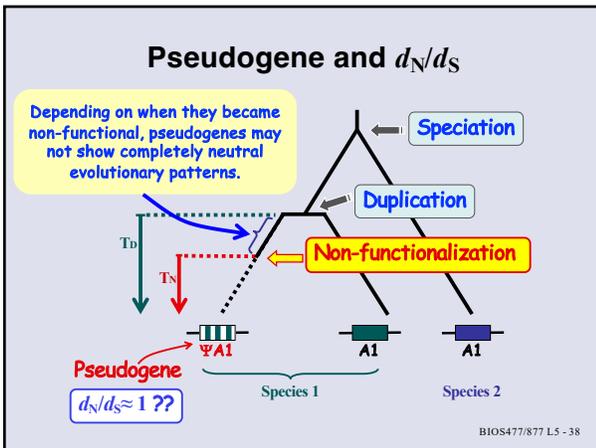
35



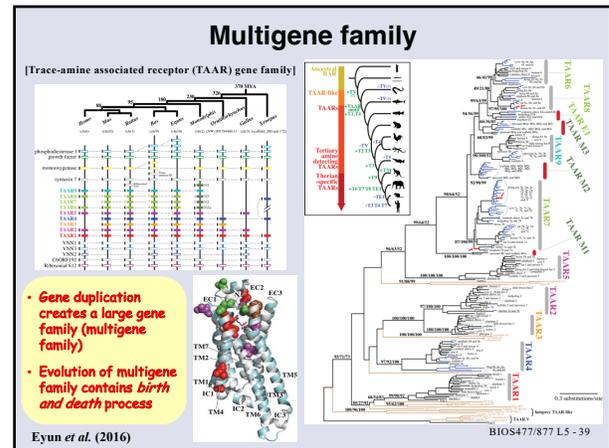
36



37



38



39

Homology vs. Similarity

- **Similarity**: the extent to which sequences are related.
→ makes no statement about descent from a common ancestor
- **Homology**: sequence similarity that can be attributed to **descent from a common ancestor**

Homology ≠ Similarity !!

- Sequences can be either **homologous** or non-homologous, but not in between (e.g., you cannot say two genes are 10% homologous!)
- **Homology** is not directly measurable or observable.
- **Similarity** is a direct measurement.

Hillis and Attwood (2005) Chapter 1, page 8

BIOS477/877 L5 - 40

40