

QBQ-2507

IBI-5035

Biologia Molecular Computacional

Prof. João Carlos Setubal

Bioquímica-Instituto de Química-USP

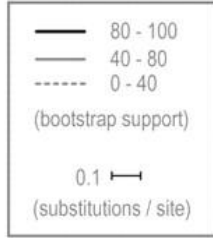
2º semestre de 2013

Preliminares

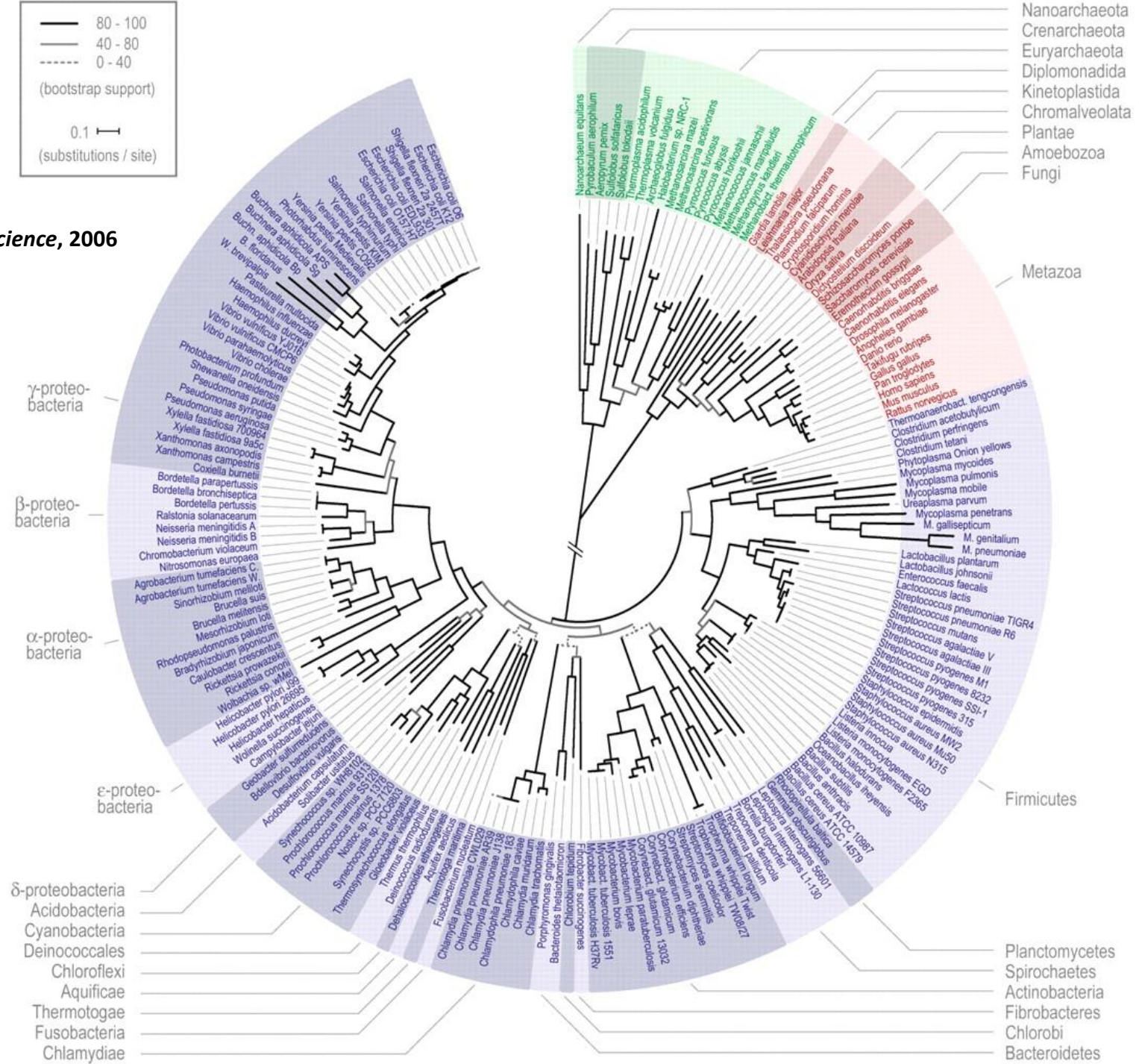
- Folha introdutória
- Descrição de possíveis projetos
- Uso de computadores
 - No lab de cada um
 - Sala multimidia do IQ (precisa reserva)
 - Sala de terminais linux no IME

Sumário do conteúdo

- Comparação de sequências
 - DNA e proteína
 - Alinhamento 2-a-2 e múltiplo
 - Sequências curtas e longas
 - Algoritmos exatos e heurísticas
- Famílias de proteínas
- Filogenia molecular
- Filogenômica
- Busca de motivos em sequências
- Transcritômica
- Metagenômica



Ciccarelli et al, *Science*, 2006



Como foi possível criar tal árvore?

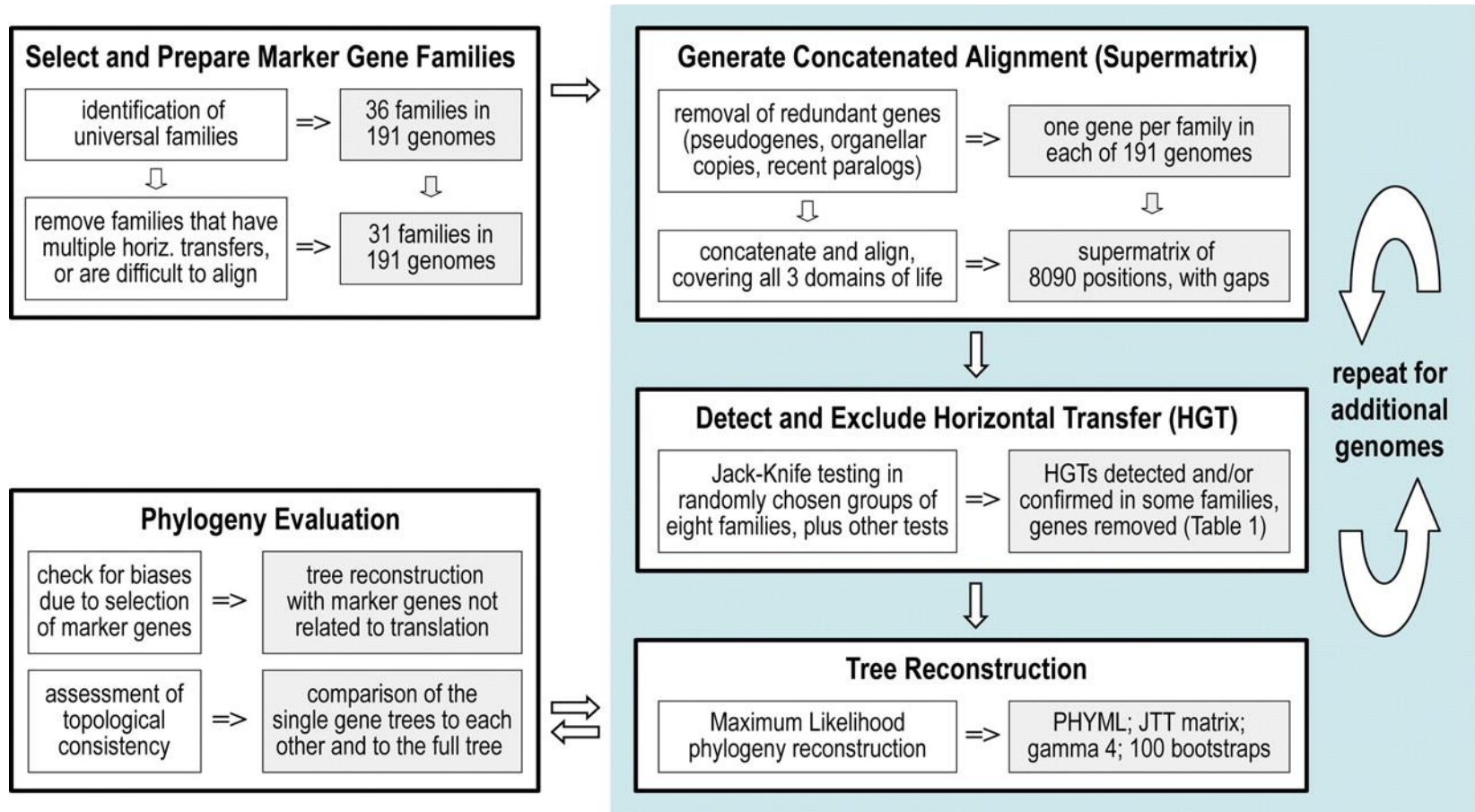
- Determinar genes que são **compartilhados** por todos os seres vivos
- O que é “compartilhar”?
 - Gene x_1 em organismo A tem função α
 - Gene x_2 em organismo B é **homólogo** e tem a **mesma função α**
 - Então x_1 e x_2 são “o mesmo gene x ” e portanto ele é compartilhado por A e B
- Quais genes são esses?

Orthologous Group	Av. Length	Annotation	Genes in Prok.	Genes in Euk.	Total Genes
COG0012	380	Predicted GTPase, probable translation factor	171	30	201
COG0016	423	Phenylalanine-tRNA synthetase alpha subunit	168	42	210
COG0018†	548	Arginyl-tRNA synthetase	175	45	220
COG0048	137	Ribosomal protein S12	168	48	216
COG0049	182	Ribosomal protein S7	169	41	210
COG0052	240	Ribosomal protein S2	168	79	247
COG0060*	956	Isoleucyl-tRNA synthetase	172	42	214
COG0080	154	Ribosomal protein L11	170	61	231
COG0081	230	Ribosomal protein L1	168	61	229
COG0085†	1138	DNA-directed RNA polymerase, beta subunit	178	60	238
COG0087	288	Ribosomal protein L3	168	54	222
COG0091	157	Ribosomal protein L22	168	75	243
COG0092	240	Ribosomal protein S3	168	30	198
COG0093	130	Ribosomal protein L14	168	41	209
COG0094	182	Ribosomal protein L5	169	36	205
COG0096	131	Ribosomal protein S8	168	55	223
COG0097	177	Ribosomal protein L6P/L9E	168	65	233
COG0098	220	Ribosomal protein S5	168	110	278
COG0099‡	133	Ribosomal protein S13	168	49	217
COG0100	145	Ribosomal protein S11	169	51	220
COG0102	167	Ribosomal protein L13	168	54	222
COG0103	172	Ribosomal protein S9	168	52	220
COG0124*	472	Histidyl-tRNA synthetase	178	31	209
COG0143*†	646	Methionyl-tRNA synthetase	180	35	215
COG0172	442	Seryl-tRNA synthetase	177	37	214
COG0184	154	Ribosomal protein S15P/S13E	168	41	209
COG0186	122	Ribosomal protein S17	170	46	216
COG0197	175	Ribosomal protein L16/L10E	168	54	222
COG0200	166	Ribosomal protein L15	168	70	238
COG0201	445	Preprotein translocase subunit SecY	178	37	215
COG0202	323	DNA-directed RNA polymerase, alpha subunit	171	45	216
COG0256	178	Ribosomal protein L18	168	50	218
COG0495	854	Leucyl-tRNA synthetase	172	43	215
COG0522	199	Ribosomal protein S4 and related proteins	174	46	220
COG0525*‡	880	Valyl-tRNA synthetase	169	37	206
COG0533	375	Metal-dependent proteases with chaperone activity	168	35	203

Questões

- Onde podemos achar as sequências dos genes?
- Como determinar compartilhamento?
- Como preparar esses dados para construir uma árvore?
- Como construir uma árvore?
- Como saber se ela está correta?
- Algo novo desde 2006?

Fig. 1. Overview of the procedure.



Respostas

- Onde podemos achar as sequências dos genes?
 - Bancos de dados públicos (NCBI)
 - BLAST
- Como determinar compartilhamento?
 - Através de comparação de sequências
- Como preparar esses dados para construir uma árvore?
 - Alinhamento múltiplo concatenado
- Como construir uma árvore?
 - Métodos de reconstrução filogenética
- Como saber se ela esta correta?
 - Inferência é um termo melhor do que construção
 - Argumentos probabilísticos
 - Transferência Horizontal (Lateral) de Genes
- Algo novo desde 2006?
 - Next Generation Sequencing (NGS)
 - Comparação de genomas completos

Outros temas

- Busca de motivos em sequências
- Transcritômica (NGS)
- Metagenômica (NGS)
- **Temas que não serão abordados**
 - Montagem de genomas
 - Redes regulatórias
 - Redes gênicas
 - Redes de interação proteína-proteína
 - Proteômica
 - Predição de estruturas
 - Biologia de sistemas
 - Como integrar dados de diferentes fontes

Cadê a informática?

- Programação
 - Perl/Python
- Algoritmos
 - Programação dinâmica
- Programas para análise
 - BLAST
 - Muscle (AM)
 - phyML (árvores)
 - MUMmer
 - Bowtie (mapeamento)