



LNCC/MCT
Petrópolis, RJ, Brasil
4 a 15 de julho de 2011

CBAB 2011

Ferramentas de Bioinformática Aplicadas à Análise de Sequências Genômicas, Metagenômicas e Transcriptômicas

Curso
CBAB/CABBIO
2011

**Ferramentas de
Bioinformática
Aplicadas à
Análise de Sequências
Genômicas,
Metagenômicas e
Transcriptômicas**

LNCC/MCT <www.lncc.br>
Petrópolis, RJ, Brasil
4 a 15 de julho de 2011

GENOMAS DE RIZÓBIOS

Mariangela Hungria

Embrapa Soja, Cx. Postal 231, 86001-970. Londrina, Paraná, Brasil

Os procariotos se estabeleceram na Terra bilhões de anos antes de qualquer célula de eucarioto, e o longo período de adaptação a diversos ambientes explica a enorme diversidade e versatilidade desses organismos.

Nesta última década, o número de genomas de procariotos sequenciados cresceu exponencialmente, e novos conhecimentos vêm sendo acumulados pelos resultados obtidos no sequenciamento de 1.650 procariotos; além disso, novas informações devem surgir a partir dos quase 6.000 sequenciamentos que estão em progresso (GOLD, 2011). A diversidade e a versatilidade que vêm sendo reveladas nesses genomas de procariotos são surpreendentes em vários aspectos, com variabilidade em tamanho, número, densidade e organização dos genes em *operons* e, de modo surpreendente, cada novo genoma tem fornecido informações intrigantes e, às vezes, inesperadas (Konstantinidis & Tiedje, 2005).

Em relação ao tamanho, os menores genomas de procariotos sequenciados incluem, nas arqueobactérias, um cromossomo com 490.885 pares de bases (pb) na termófila *Nanoarchaeum equitans* estirpe Kin4-M, simbiote obrigatório de outra arqueobactéria, *Ignicoccus* sp. e, nas bactérias, um cromossomo com 615.980 pb de *Buchnera aphidicola* estirpe Bp, também um simbiote intracelular do afídeo *Baizongia pistacea*. Esses pequenos genomas despertam o interesse pelo conhecimento de qual seria o conjunto mínimo de genes e de processos metabólicos essenciais para a sobrevivência de um organismo vivo. Já os maiores genomas sequenciados incluem *Burkholderia xenovorans* estirpe LB400, com 9.767.620 pb distribuídos em três cromossomos e o único cromossomo com 9.105.828 pb da estirpe USDA 110 de *Bradyrhizobium japonicum*. Surpreendentemente, mesmo na mesma espécie, diferenças marcantes entre estirpes podem ser encontradas, por exemplo, o genoma de *Escherichia coli* estirpe K12 foi estimado em 4.646.332 pb e 4.337 genes foram identificados, enquanto que na estirpe O157:H7 o genoma de 5.498.450 bp codifica 5.449 genes. Outro exemplo é o da bactéria fotossintética *Bradyrhizobium* sp. estirpe BTAi1, com mais de 1 milhão de pb do que a estirpe ORS278 (GOLD, 2011)

Contudo, mesmo com o crescimento exponencial no número de genomas sequenciados, o número de novas famílias de proteínas continua incrementando regularmente a cada novo genoma publicado, particularmente entre procariotos. Modelos matemáticos predizem, inclusive, que novos genes serão descobertos mesmo após o sequenciamento de centenas de genomas por espécie e podem conduzir à descoberta de inúmeras novas famílias de proteínas com propriedades bioquímicas de interesse. Além disso, o número ainda elevado de CDSs (*coding DNA sequences*) com função hipotética ou hipotética conservada continua a crescer, particularmente entre procariotos, indicando que o nosso conhecimento sobre sequências de nucleotídeos ainda é

pequeno. Mesmo em relação aos genes conhecidos e organizados em *operons*, os genomas apresentam grande variação na ordenação desses genes, com poucos *operons* sendo mantidos entre linhagens evolutivamente distantes e existem diferenças marcantes na estabilidade dos mesmos.

Finalmente, cabe salientar que a grande maioria dos genomas de bactérias atualmente disponível, ou em progresso, é de patógenos humanos, ou de patógenos de animais, ou de plantas. Relativamente poucos são os genomas de bactérias de vida livre e não-patogênicas e apenas uma porcentagem mínima inclui representantes da biodiversidade dos trópicos

O processo de fixação biológica do nitrogênio atmosférico (N_2) é realizado por alguns microrganismos cujo principal hábitat é o solo. Embora o N_2 constitua, aproximadamente, 80% dos gases atmosféricos, nenhum animal ou planta consegue utilizá-lo como nutriente, devido à tripla ligação que existe entre os dois átomos do N_2 , uma das mais fortes na natureza. Contudo, os gases atmosféricos também se difundem para o espaço poroso do solo, e o N_2 consegue ser aproveitado por alguns microrganismos procariotos (arqueobactérias e, principalmente, bactérias) que ali habitam, graças à ação de um complexo enzimático denominado dinitrogenase, capaz de romper a tripla ligação do N_2 e reduzi-lo a amônia (NH_3), a mesma forma obtida no processo industrial. Essas bactérias, também denominadas como diazotróficas ou fixadoras de N_2 , se associam a diversas espécies de plantas em diferentes graus de especificidade. A maior contribuição e, por isso, a mais estudada, ocorre com bactérias que se associam simbioticamente com diversas plantas da família *Leguminosae*, conhecidas como leguminosas. As bactérias que se associam a essas leguminosas são chamadas, coletivamente, de rizóbios.

A distribuição de rizóbios em uma variedade de bactérias não relacionadas filogeneticamente tem despertado a atenção e levantado diversas discussões sobre a evolução do processo de fixação biológica do N_2 . Uma hipótese é a de que o N—particularmente amônia—ficou limitante ao crescimento microbiano, conduzindo à evolução da fixação biológica do N_2 . Evidências para essa teoria residem no ponto comum entre as bactérias diazotróficas, o complexo da enzima nitrogenase, que é codificado por uma família de genes que, aparentemente, evoluíram a partir de uma unidade altamente conservada, embora tenham ocorrido contribuições ocasionais de episódios de duplicação de genes, recrutamento, fusão e transferência de genes. Os genes de nodulação tiveram uma origem mais recente, surgindo com a evolução das leguminosas hospedeiras, de modo que a filogenia observada com os genes ribossomais em geral não é congruente com aquela obtida com os genes de nodulação (Lloret & Martínez-Romero, 2005). Comparações interessantes também surgem pela análise de genomas de bactérias diazotróficas e bactérias patogênicas relacionadas filogeneticamente, com a constatação de um grupo comum de genes (Carvalho *et al.*, 2010).

Apesar da importância dos diazotróficos simbióticos na ciclagem global do nitrogênio, relativamente poucos genomas completos estão disponíveis. O primeiro relato foi do sequenciamento do plasmídeo simbiótico de *Rhizobium* sp. estirpe NGR234, com a identificação de 416 genes funcionais, incluindo 26 genes para nodulação e 26 genes para fixação do nitrogênio (Freiberg *et al.*, 1997). A seguir, genomas completos foram descritos para *Mesorhizobium loti* MAFF303099 (Kaneko *et al.*, 2000), *Sinorhizobium (=Ensifer) meliloti* 1021 (Galibert *et al.*, 2001), *B. japonicum* USDA 110 (Kaneko *et al.*, 2002), *Rhizobium etli* bv. phaseoli CFN 42 (González *et al.*,

2006), *R. leguminosarum* biovar *viciae* 3841 (Young *et al.*, 2006), *Bradyrhizobium* sp. ORS278 e BTAi1 (Giraud *et al.*, 2007), *Azorhizobium caulinodans* ORS571 (Lee *et al.*, 2008), *Cupriavidus taiwanensis* LMG19424 (Amadou *et al.*, 2008). Esses genomas de rizóbios são compostos por um cromossomo (exceto *Cupriavidus*, com dois cromossomos) e todos, exceto *B. japonicum* estirpe USDA 110, *Bradyrhizobium* sp. ORS278 e *A. caulinodans* ORS571, carregam até seis plasmídeos, variando de 100 kb a 2 Mb. Os genes relacionados à simbiose estão localizados em plasmídeos, ou em ilhas no cromossomo, provavelmente adquiridas por transferência horizontal de genes, embora a natureza precisa desses genes ainda não tenha sido completamente elucidada. Outras informações de genomas estão disponíveis na Tabela 1.

O Brasil é, provavelmente, o país que mais tem utilizado os benefícios do processo de fixação biológica do nitrogênio em culturas de grãos, por isso, torna-se importante desenvolver estudos de genômica com estirpes brasileiras, particularmente aquelas utilizadas em inoculantes comerciais, mas o custo do sequenciamento de um genoma completo, embora tenha caído drasticamente, ainda é bastante elevado. Uma estratégia útil, particularmente antes de surgir o pirosequenciamento, foi denominada de panorama genômico, e consiste no sequenciamento parcial do genoma, que foi utilizado em *Rhizobium* sp. estirpe NGR234, curada do plasmídeo simbiótico pNGR234a (Viprey *et al.*, 2000). A leitura de 2.275 sequências nessa bactéria resultou em mais de 575 kb, correspondendo a, aproximadamente, 10% do genoma total. Ainda nesse trabalho, a busca de homologia em bancos de dados resultou na identificação de 1.130 sequências codificando proteínas putativas e a análise dessas sequências, em comparação com outros genomas, sugere que esse número foi suficiente para representar o genoma. Além disso, diversas sequências pioneiras foram identificadas, indicando que informações relevantes podem ser obtidas, a um custo acessível, pelo sequenciamento parcial (Viprey *et al.*, 2000).

Essa estratégia foi utilizada com sucesso por nosso grupo, para obter o panorama genômico da estirpe CPAC 15 (=SEMIA 5079) de *B. japonicum* (Godoy *et al.*, 2008). Um total de 4.328 leituras de bibliotecas do tipo *shotgun* resultaram no depósito de 2.046.740 pb com um phred score ≥ 20 , em uma cobertura real de cerca de 13% do genoma, com um total de 1.371 CDSs anotadas, correspondendo a $\approx 16\%$ das proteínas estimadas, sendo 729 classificadas como válidas, 312 como hipotéticas conservadas e 330 como hipotéticas. A classificação funcional das CDSs pelo banco de dados COG mostrou genes putativos em quase todas as classes. Além disso, a comparação entre as estirpes de *B. japonicum* USDA 110, sequenciada em 2002 e a CPAC 15, demonstrou uma proporção muito próxima na porcentagem de CDSs encontradas nas diferentes categorias de COG, uma constatação importante para a validação da estratégia de sequenciamento parcial para o conhecimento inicial do genoma de uma bactéria. Contudo, as duas estirpes são surpreendentemente diferentes, uma vez que pelo menos 35% das CDSs da CPAC 15 apresentaram maior similaridade com outros microrganismos que não a USDA 110. Várias CDSs foram anotadas como genes putativos que codificam proteínas relacionadas à nodulação, como os genes *nodB*, *nodD2*, *nodW*, *nodV* e a proteína efetora *nopP* (sistema de secreção do tipo III) e à fixação do N_2 , como os genes *nifE*, *fixP*, *fixQ*, *fixR* e *fixO*, além de outras proteínas envolvidas no ciclo metabólico do N. Alguns desses genes putativos, como *nodW* e *nopP* apresentam possibilidade de utilização na indústria de inoculantes. No sequenciamento parcial do genoma da estirpe CPAC 15, 5,76 % das CDSs válidas codificam para enzimas que participam de quase todas as vias de biodegradação de xenobióticos, podendo representar um reservatório importante de genes com potencial biotecnológico.

Em nosso laboratório também foi obtido um panorama genômico da estirpe PRF 81 (=SEMIA 4080) de *R. tropici*, aprovada para a produção de inoculantes comerciais para a cultura do feijoeiro (Pinto *et al.*, 2009). Pela técnica de eletroforese em campo pulsado (PFGE, *pulse field gel electrophoresis*) e pelos perfis de plasmídeos o genoma da estirpe PRF 81 foi estimado em 7,85 Mb, consistindo de um genoma circular e quatro plasmídeos. Foram realizadas leituras de 9.026 clones obtidos por *shotgun*, que foram montados em 1.668 *phrap contigs* cobrindo, aproximadamente, 31% do genoma. Pela anotação foram identificados 2.135 CDSs, somente 57,2% mostrando similaridade com genes com função conhecida. O genoma contém um mosaico de genes, com CDSs mostrando maior similaridade com uma variedade de microrganismos, 134. A capacidade saprofítica e a competitividade elevadas da PRF 81 podem estar relacionadas a uma variedade de genes relacionados ao transporte, biodegradação de xenobióticos, defesa, sensoriamento ambiental, metabolismo de compostos pouco comuns e secreção de proteínas, muitas das quais nunca haviam sido relatadas antes em *R. tropici*. Novidades também foram encontradas nos genes de nodulação (*nodG*, um duplo sistema *nodII*, *nodT*, *nolF*, *nolG*) e genes de polisacarídeos capsulares (KPS). Similaridade mais elevada nos genes de nodulação e KPS foram constatadas com *Sinorhizobium* e não com outros simbiotes comuns do feijoeiro—*R. etli* e *R. leguminosarum*—sugerindo que o hospedeiro original de *R. tropici* pode ser outra leguminosa que não o feijoeiro.

Pode-se concluir, portanto, que, embora o processo de fixação biológica do nitrogênio apresente uma grande contribuição para o agronegócio brasileiro, bem como a constatação de que, pelo menos no caso da soja, o Brasil é o país que mais se beneficia do processo biológico, os estudos de genoma de rizóbios, em nosso País, ainda são inexpressivos. Atualmente, com os avanços nas técnicas de sequenciamento, a Embrapa Soja e o LNCC estão empenhados em obter o genoma completo dos rizóbios mais relevantes para o agronegócio brasileiro.

Tabela 1. Informações de genomas de diazotróficos simbióticos disponíveis.

Organismo	Kb	# crom	#plam.	GC%
<i>Azorhizobium caulinodans</i> ORS 571	5369	1	0	67.3
<i>Bradyrhizobium japonicum</i> USDA110	9106	1	0	64.1
<i>Bradyrhizobium</i> sp ORS278	7456	1	0	65.5
<i>Bradyrhizobium</i> sp BTAi1	8264	1	1	64
<i>Ensifer medicae</i> WSM419	6818	1	3	62
<i>Ensifer meliloti</i> 1021	6692	1	2	62.7
<i>Methylobacterium</i> sp. 4-46	7737	1	2	71.5
<i>Methylobacterium nodulans</i> ORS2060	7724	1	7	68
<i>Mesorhizobium loti</i> MAFF303099	7596	1	2	62.7

<i>Mesorhizobium opportunistum</i> WSM2075	6854	1	0	62
<i>Rhizobium etli</i> CFN42	6530	1	6	61.3
<i>Rhizobium etli</i> CIAT 652	6448	1	3	61.7
<i>Rhizobium leguminosarum</i> bv <i>viciae</i> 3841	7751	1	6	63
<i>Rhizobium leguminosarum</i> bv <i>trifolii</i> WSM2304	6873	1	4	61
<i>Rhizobium leguminosarum</i> bv <i>trifolii</i> WSM1325	7418	1	5	61
<i>Rhizobium</i> sp. NGR234 (ANU265)	3925	1	2	63
<i>Cupriavidus taiwanensis</i> LMG19424	6476	2	1	67
<i>Burkholderia phymatum</i> STM815	8676	2	2	62

Referências:

AMADOU, C.; PASCAL, G.; MANGENOT, S. et al. Genome sequence of the β -rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Research*, Published online DOI: 10.1101/gr.076448.108, 2008.

CARVALHO, F.M.; SOUZA, R.C.; BARCELLOS, F.G.; HUNGRIA, M.; VASCONCELOS, A.T.R. Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order Rhizobiales. *BMC Microbiology*, v.10, 37, 2010. (DOI: 10.1186/1471-2180-10-37). <http://www.biomedcentral.com/1471-2180/10/37>.

FREIBERG, C.; FELLAY, R.; BAIROCH, A.; BROUGHTON, W.J.; ROSENTHAL, A.; PERRET, X. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature*, v.387, p.394-401, 1997.

GALIBERT, F.; FINAN, T.M.; LONG, S.R.; PUHLER, A. et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, v. 293, p. 668-72, 2001.

GIRAUD, E. ; MOULIN, L.; VALLENET, D. et al. Legumes symbioses: absence of nod genes in photosynthetic bradyrhizobia. *Science*, v.316, p.1307-1312, 2007.

GODOY, L.P.; VASCONCELOS, A.T.R.; CHUEIRE, L.M.O.; SOUZA, R.C.; NICOLÁS, M.F.; BARCELLOS, F.G.; HUNGRIA, M. Genomic panorama of *Bradyrhizobium japonicum* CPAC 15, a commercial inoculant strain largely established in Brazilian soils and belonging to the same serogroup as USDA 123. *Soil Biology & Biochemistry*, v.40, p.2742-2753, 2008.

GOLD[™] (Genomes OnLine Database). Genomes OnLine Database. (online) <http://www.genomesonline.org/>. (acesso em 20 de junho de 2011).

GONZÁLEZ, V.; BUSTOS, P.; RAMIREZ-ROMERO, M.A.; MEDRANO-SOTO, A. et al. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biology*, v.4, n.6, R36, 2003. KANEKO, T.; NAKAMURA, Y.; SATO,

S.; MINAMISAWA, K. et al. Complete genome sequence of 8) the nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA 110. *DNA Research*, v.9, p.189-197, 2002a.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; MINAMISAWA, K. et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research*, v.9, supplement, p.225-256, 2002b.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; ASAMIZU, E. et al. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research*, v.7, p.331-338, 2000.

KONSTANTINIDIS, K.T.; TIEDJE, J.M. Genomic insights that advance the species definition for prokaryotes. *Proceeding of the National Academy of Sciences of U.S.A.*, v.102, p.2567-2572, 2005.

LEE, K.B.; DE BACKER, P.; AONO, T. et al. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571. *BMC Genome*, v.9, p.271, 2008.

LLORET, L.; MARTÍNEZ-ROMERO, E. Evolución y filogenia de *Rhizobium*. *Revista Latinoamericana de Microbiología*, v.47, p.43-60, 2005.

PINTO, F.G.S.; CHUEIRE, L.M.O.; VASCONCELOS, A.T.R.; NICOLÁS, M.F.; ALMEIDA, L.G.P.; SOUZA, R.C.; MENNA, P.; BARCELLOS, F.G.; MEGÍAS, M.; HUNGRIA, M. Novel genes related to nodulation, secretion systems, and surface structures revealed by a genome draft of *Rhizobium tropici* strain PRF 81. *Functional and Integrative Genomics*, v.9, n.2, p.263-270, 2009.

VIPREY, V.; ROSENTHAL, A.; BROUGHTON, W.K.; PERRET, X. Genetic snapshots of the *Rhizobium* species NGR234 genome. *Genome Biology*, v.1, p.1-17, 2000. (<<http://genomebiology.com/2000/1/6/research/0014.1>>).

YOUNG, J.P.W.; CROSSMAN, L.C.; JOHNSTON, A.W.B., et al. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology*, v.7, R34, 2006.