Our research
– your profit

Pan-*Thermus*

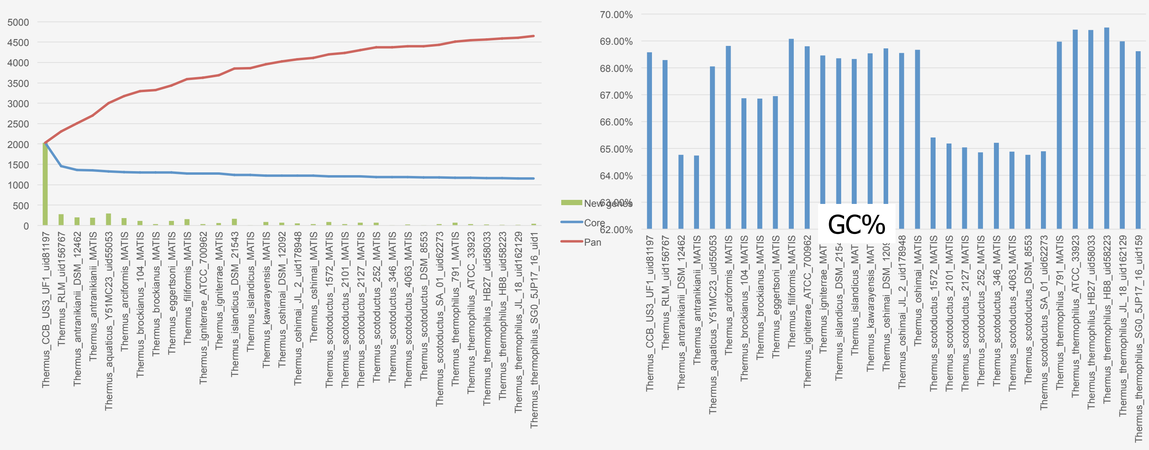
Sigmar K. Stefánsson[1,2], Snædís Björnsdóttir[1] , Sólveig Pétursdóttir[1] , Ólafur H. Friðjónsson[1], Guðmundur Ó. Hreggviðsson[1,2]

[1] Matís ltd. Vínlandsleið 12, 113 Reykjavík, Iceland; [2] Faculty of Life and Environmental Sciences, University of Iceland, Sturlugata 7, 101 Reykjavík Iceland

Introduction

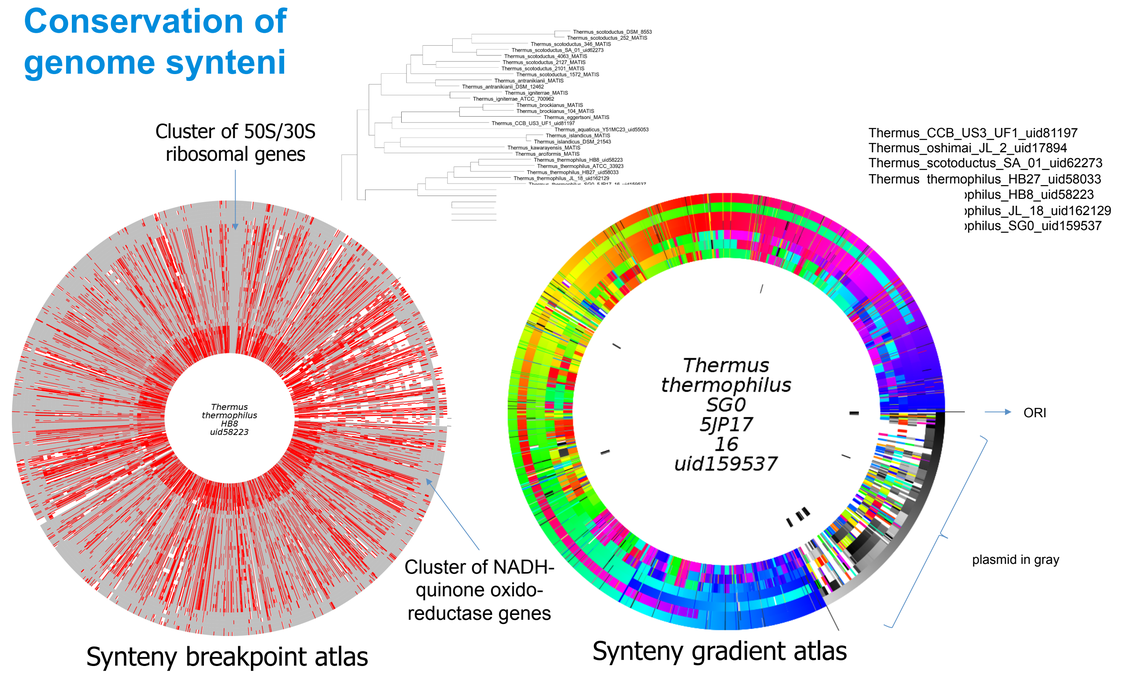
Bacteria of the genus *Thermus* are common in geothermal habitats worldwide. Great phylogenetic diversity is observed for *Thermus* bacteria in Icelandic hot springs, although few phenotypically distinguishing characters have been observed. This indicates that adaptive traits are subtle physiological differences in relation to physicochemical conditions. The aim of the Pan-*Thermus* project is to resolve ecological adaptations and evolutionary mechanisms contributing to speciation within the genus *Thermus.* Our approach involves both the mapping of the distribution of *Thermus* species in relation to environmental gradients and defining the peripheral genes and the core genome of the genus. High-resolution data on the distribution of 16S rRNA genes in water-, sediment- and biomat-samples from 32 diverse hot springs has been acquired using 454 sequencing. In addition, the genomes of 17 *Thermus* strains have been sequenced at Matís. The genomes are being analysed along with 15 published genome sequences. Software for pangenomic analysis, the SIMSNavigator, has been developed for the project. The SIMSNavigator allows genome comparisons, analysis and visualization of pan- and core genomes, gene order, genetic islands and phylogenetics. It is linked with the iPath v2 software for construction of metabolic pathways.

Genome comparison

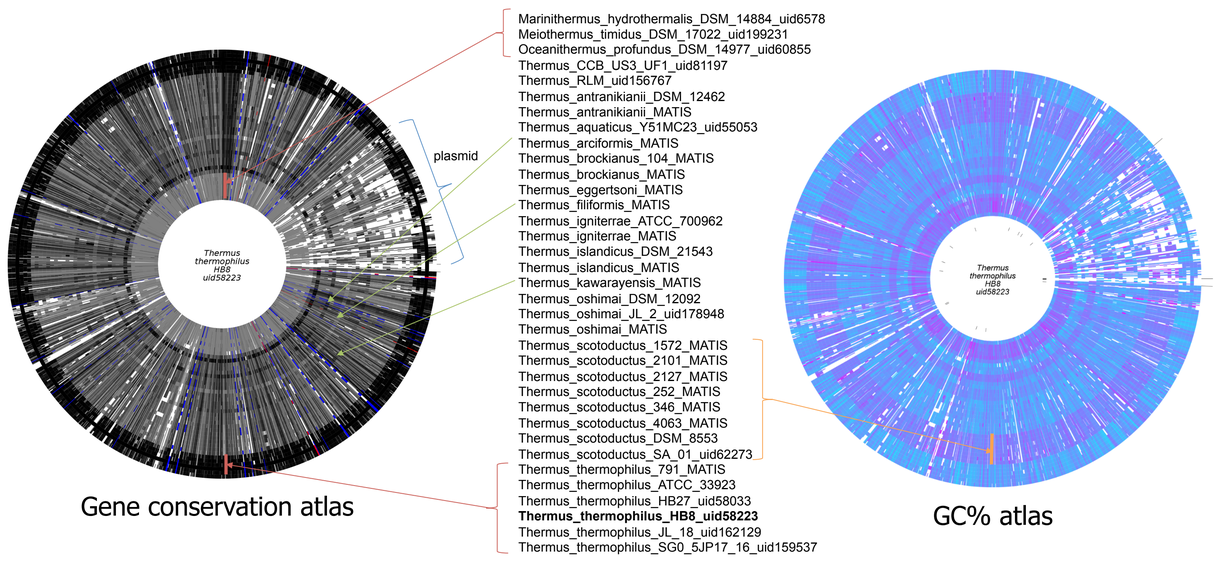


The graph on the left is a pan-core genome plot for 32 fully sequenced *Thermus* strains. The plot shows the increased number of genes (y-axis) in the pan genome with each strain added (x-axis) and a concomitant decrease in the core genome. The columns show the number of genes that are added to the pan genome with each strain. The graph on the right shows differences in the average GC% (y-axis) for the 32 strains (x-axis). Lower GC% values are observed for the *T. scotoductus* strains than for the *T. thermophilus* strains. These species exhibit the lowest and highest growth temperatures, respectively, of *Thermus*.

Conservation of genome synteni

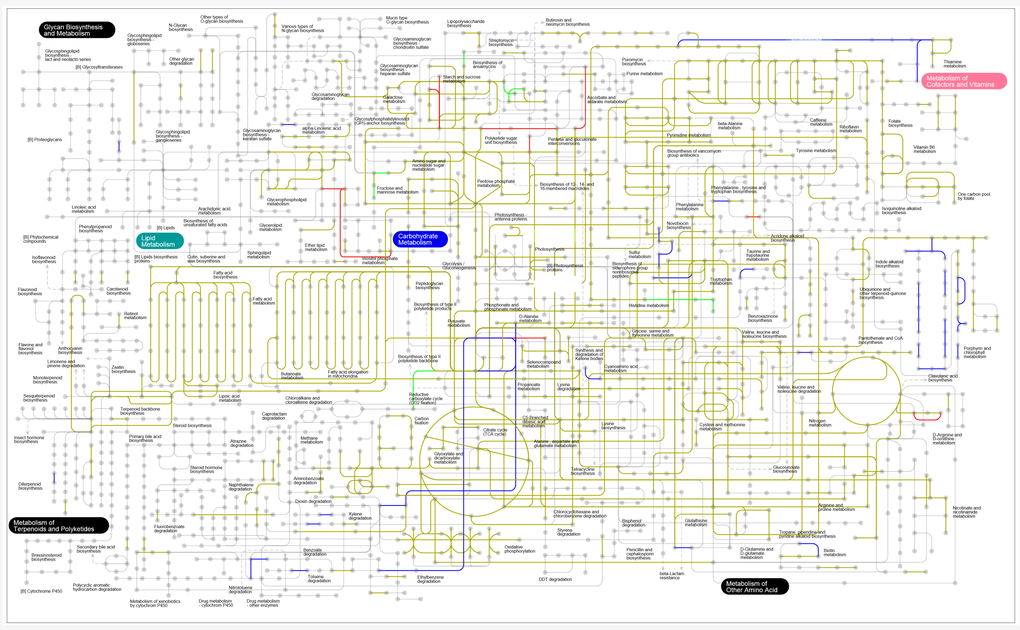


The genetic atlases above portray two different ways of visualizing genome synteny. The atlas on the left shows in red where breaks occur in the synteny compared to the reference strain, *T. thermophilus* HB8. The phylogenetic tree between the atlases is constructed from similarity based on synteny. The atlas on the right shows shuffling of gap-closed *T. thermophilus* genomes compared to the reference strain, *T. thermophilus* SG0. Colour gradients are shown, initiating from the origin of replication (*ori*). The highest conservation of synteny is usually found around the *ori*. The large inversions in two of the *T. thermophilus* genomes are flanked by identical genes, which might indicate assembly errors.



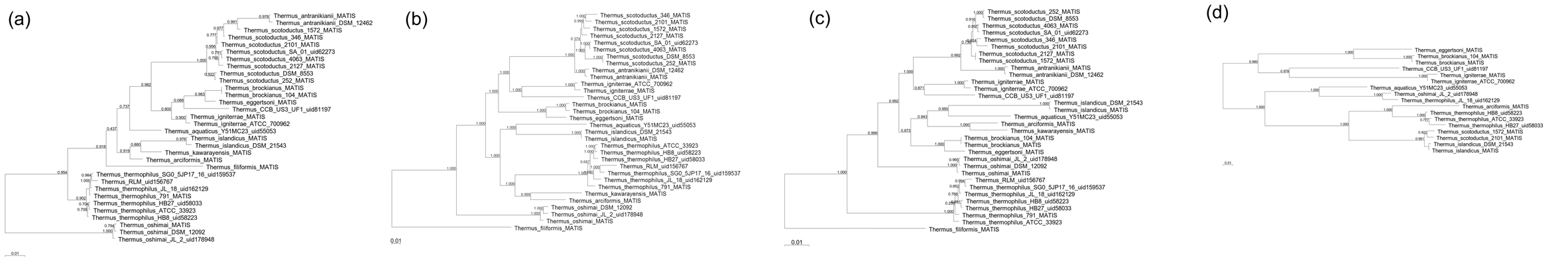
The genetic atlases above are generated with SIMSNav. The atlas on the left shows which genes in *T. thermophilus* HB8 are found in other genomes and their conservation. Black colour indicates high gene conservation. The atlas on the right shows differences in GC% between species and correlates with the graph above. Relationship between species, such as conservation between *T. thermophilus* and *T. arciformis* and *T. kawarayensis* is apparent that is not revealed in the phylogenetic trees based on ribosomal genes shown below.

Metabolic pathways



The map above compares metabolic pathways in three *Thermus* genomes. Pathways that are exclusively found in *T. oshimai, T. islandicus* and *T. scotoductus* 1572 are displayed in blue, green and red, respectively. Proteins involved in the remaining coloured pathways are encoded for in the *Thermus* core genome.

Phylogeny of Thermus species



Fully sequenced genomes of numerous *Thermus* species allow different approaches for resolving their phylogeny. Tree a) is a classical phylogenetic tree based on 16S rRNA gene sequences. Tree b) is based on the core genes found in all species except for genes encoding rRNAs and ribosomal proteins. Tree c) is based on a cluster of 34 highly conserved ribosomal genes. Their synteny is highly conserved as is evident from the synteny atlas. Some genes are sporadically distributed among *Thermus* species and strains and could be of importance for adaptation and speciation. Genes for biosynthesis of the compatible solute mannosyl glycerate are such an example. Mannosyl glycerate protects against both heat and salt stress and these genes are found exclusively in *T. thermophilus,* which exhibits the highest growth temperature and is the only *Thermus* species that thrives in marine environments. Tree d) is based on genes for which a phylogeographic distribution pattern is observed, among them are genes encoding UV damage endonucleases.