HAL
open science

TopoIBase: a comprehensive database dedicated to type IA DNA-topoisomerases

Nicolas Daccord, Eduardo Corel, Damien Correia, Anaïs Louis, Hélène Débat, Vladimir Daric, Marc Nadal, Claudine Devauchelle, Franck F. Samson

To cite this version:

Nicolas Daccord, Eduardo Corel, Damien Correia, Anaïs Louis, Hélène Débat, et al.. TopoIBase: a comprehensive database dedicated to type IA DNA-topoisomerases. JOBIM 2015 - Journées Ouvertes Biologie Informatique Mathématiques, Jul 2015, clermont-ferrand, France. , 2015. hal-02801275

HAL Id: hal-02801275

https://hal.inrae.fr/hal-02801275

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TopolBase : a comprehensive database dedicated to type IA DNA-topoisomerases.

Nicolas DACCORD[1], Eduardo COREL[1,2], Damien CORREIA[1,4], Anais LOUIS[1], Hélène DEBAT[3,4,5], Vladimir DARIC[4], Marc NADAL[4,5], Claudine DEVAUCHELLE[1] and Franck SAMSON[1]

[1]LaMMe, UMR8071 CNRS, 23 bvd de France, Université d'Evry-Val d'Essonne IBGBI, 91037 Evry
[2]IBPS, UMR 7138, Université Pierre et Marie Curie, 7 quai Saint-Bernard, 75252 Paris
[3]Université Versailles St-Quentin, 45 avenue des Etats-Unis, 78035 Versailles
[4]IGM, UMR 8621 CNRS, Université Paris-Sud, Bât. 409, 91405 Orsay Cedex
[5]Institut Jacques Monod, UMR 7592 CNRS, Université Paris Diderot

Topoisomerases

- Proteins discovered in 1971 (Wang et al., 1971)
- Function : stabilization of the DNA topology

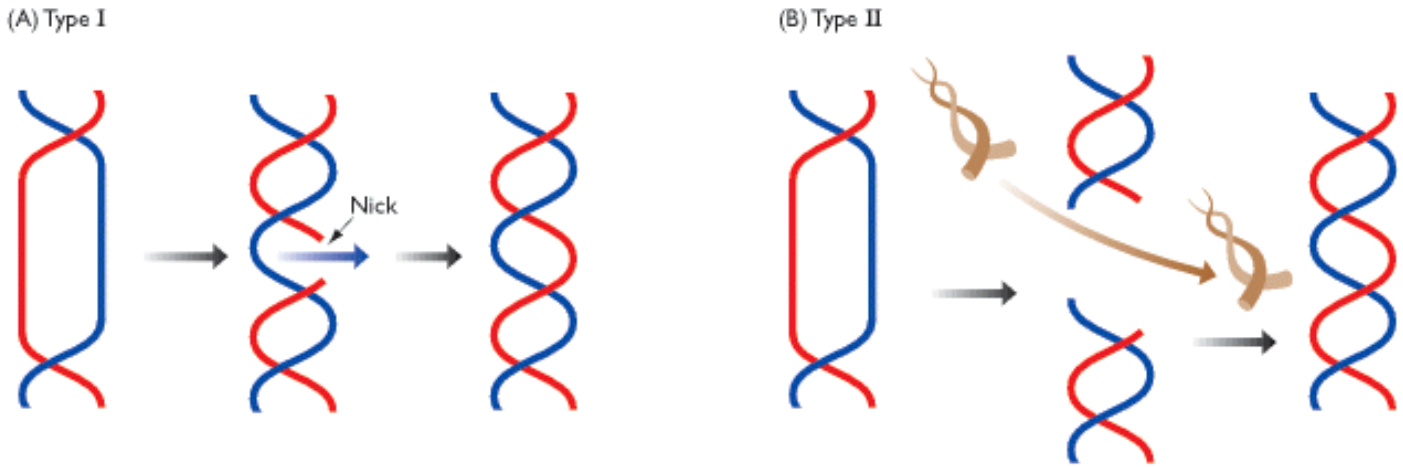


Figure: Mechanisms of action of type I et II topoisomerases.

Diversity

- Many types et sub-types of differents origins
- Different activities and dependancies
- Topoisomerases IA.

Topoisomerases			Caracteristics			Activities			
Types	Sub-types	Name	Mg2+	ATP	Cut	Releases negative supercoiling	Releases positive supercoiling	Insets negative supercoiling	Insets positive supercoiling
I	IA	I	+	-	Simple strand	+	-	-	-
		TopR	+	+	Simple strand	+	-	-	+
	IB	TopoIB	-	-	Simple strand	+	+	-	-
II	IIA (bact.)	Gyrase	+	+	Double strand	-	+	+	-
	IIA (euca.)	Topo2	+	+	Double strand	+	+	-	-
	IIB (arch.)	TopoIV	+	+	Double strand	+	+	-	-

Figure: Types and sub-types of topoisomerases and enzymatic activity variations.

Objectives

- Topoisomerases IA are difficult to study
- Lack of protein database dedicated to topoisomerases
- Objective : develop an expertised database dedicated to topoisomerases
- Expertise : data selection, VLD (Variable Length Decoding : Didier et al., 2012 [1]) classification and annotations

[1] Didier, G., Corel, E., Laprevotte, I., Grossmann, A., and Devauchelle, C. (2012). Variable length local decoding and alignment free sequence comparison.
Theoretical Computer Science, 462:1–11

Model

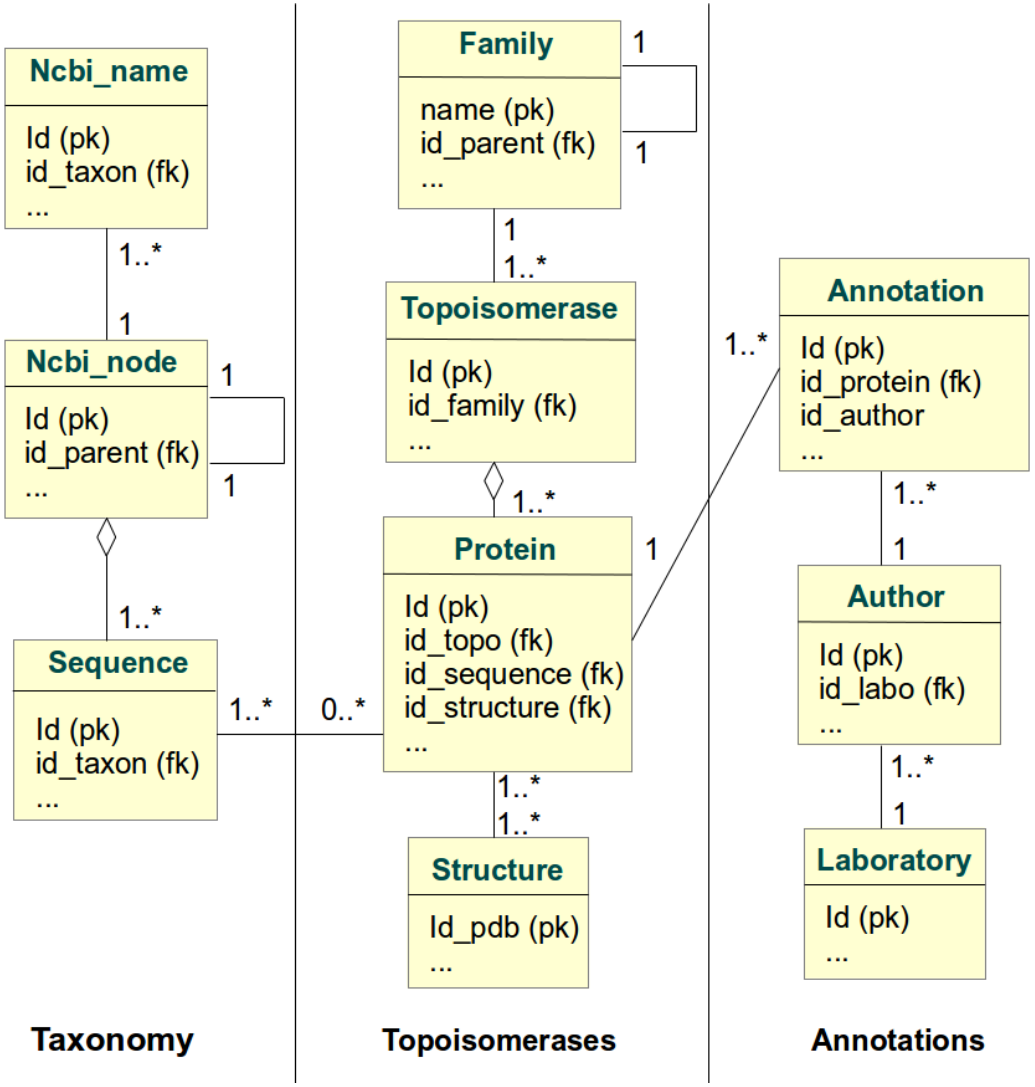


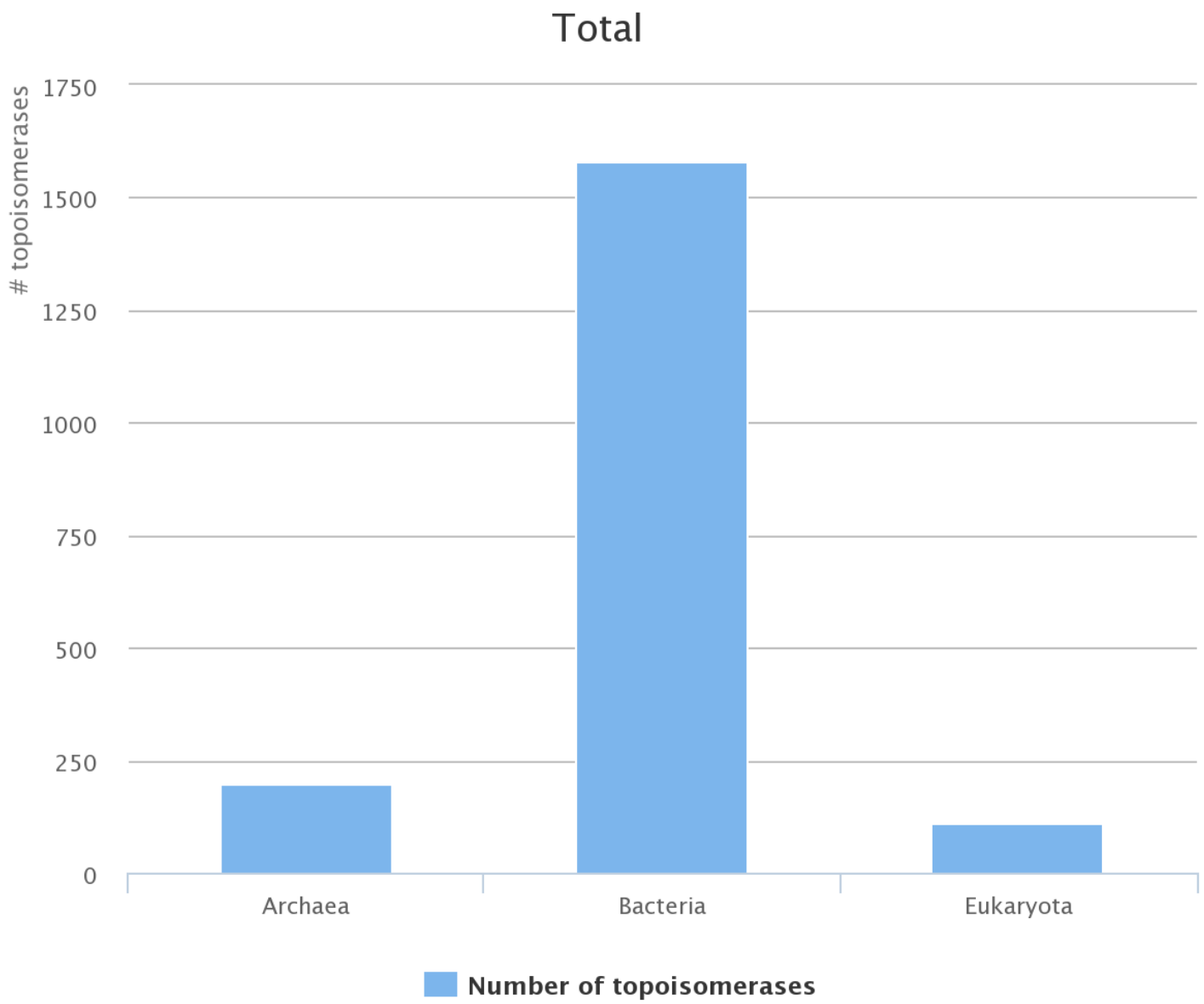
Figure: Simplified UML diagram of the database.

Data selection and expertise

- Topoisomerases : TBLASTN on fully sequenced genomes
- Classification : VLD (Didier et al. 2012 [1]) + hierarchical ascendant clustering
- Annotations by experts

[1] Didier, G., Corel, E., Laprevotte, I., Grossmann, A., and Devauchelle, C. (2012). Variable length local decoding and alignment free sequence comparison.
Theoretical Computer Science, 462:1–11

Database statistics



Database statistics (2)

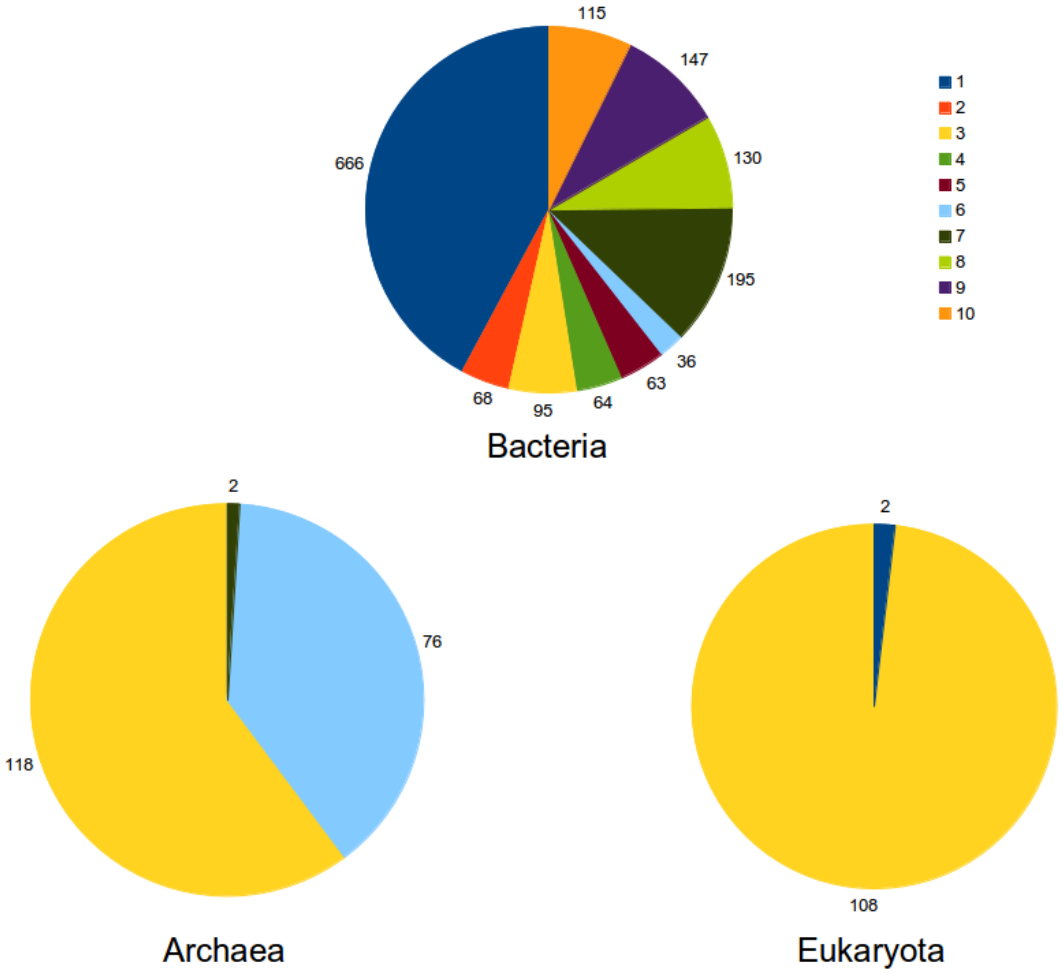


Figure: Topoisomerases distribution in domains.

Graphical interface

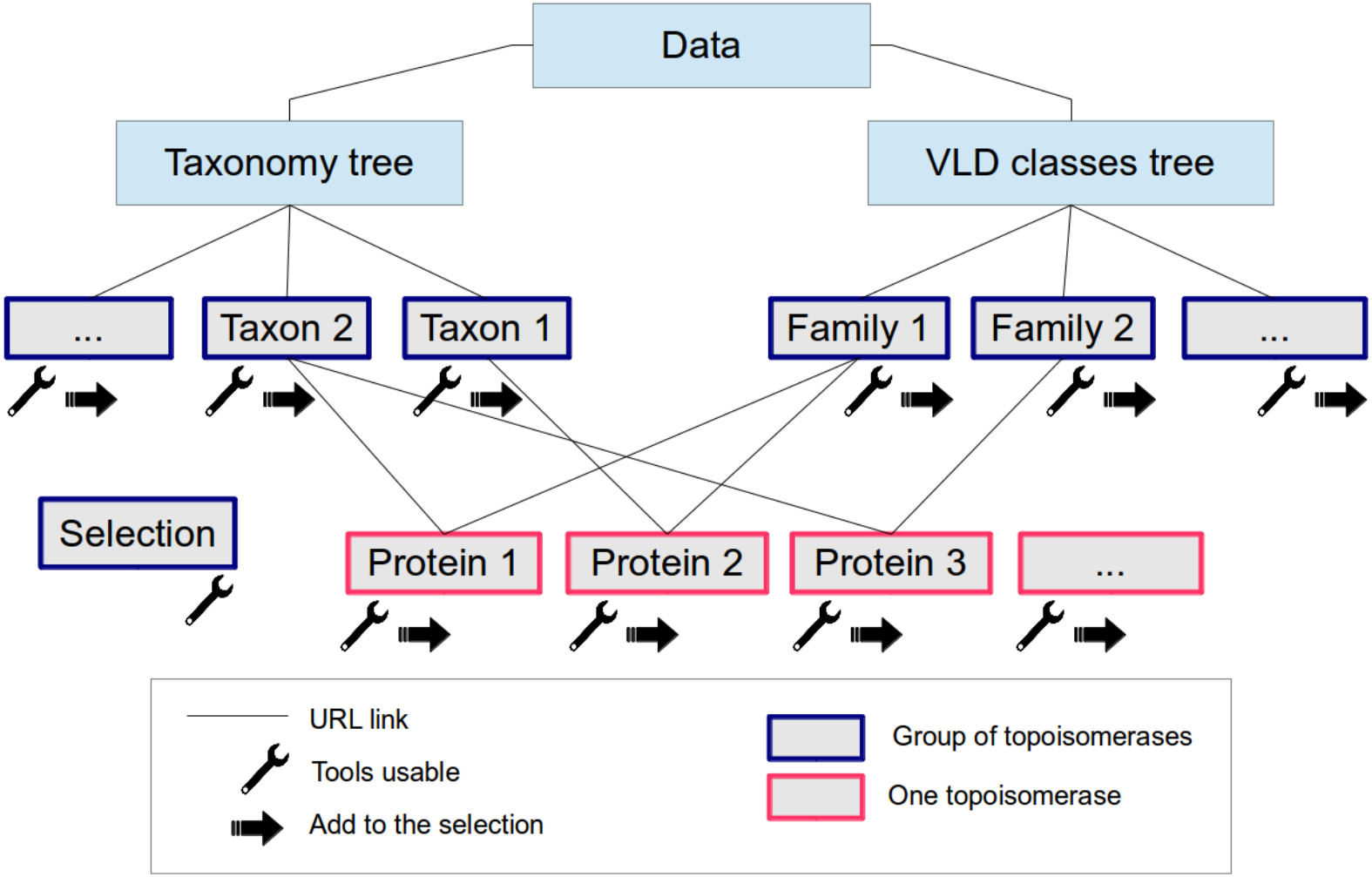


Figure: Data browsing on the interface.

Taxonomy

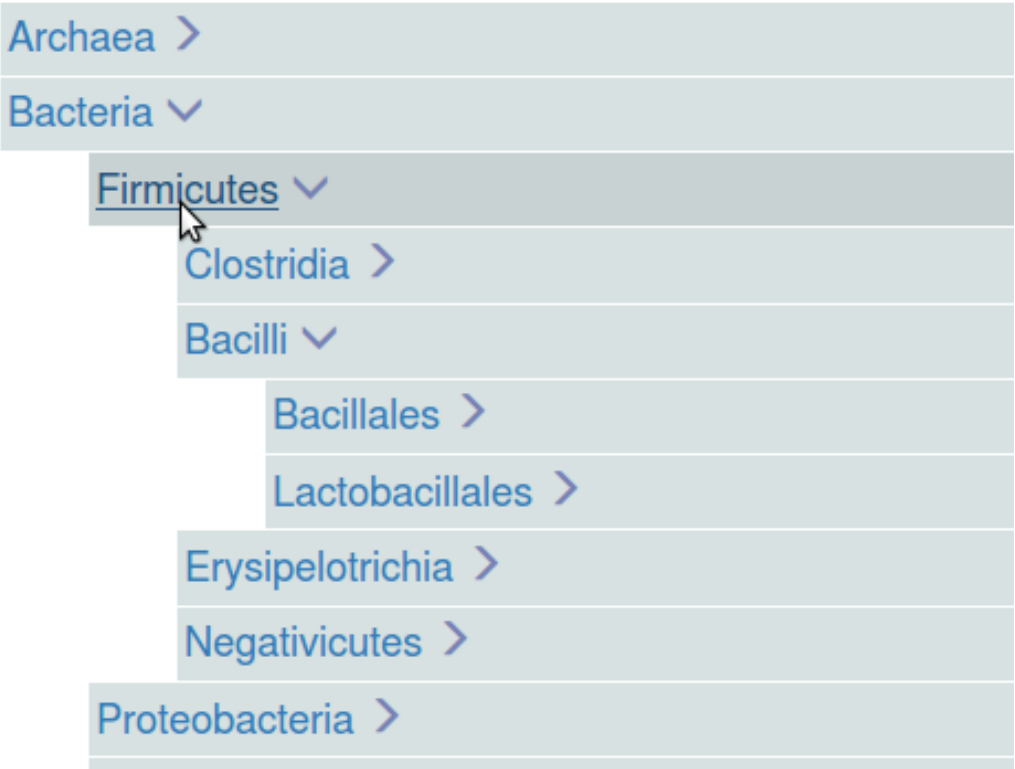


Figure: Part of the taxonomy tree.

Taxonomy

Download | Blast | Muscle | VLD | Add to selection

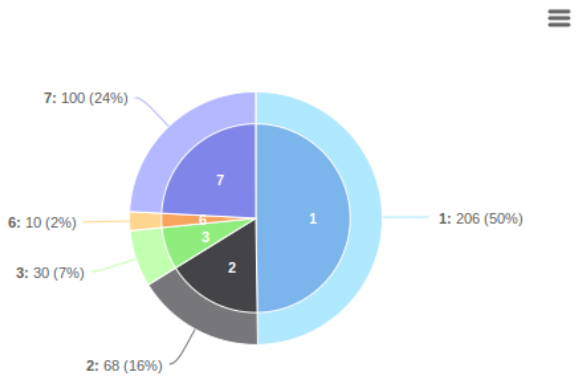
Taxon : Firmicutes

Select all ☑	Name	Family	Sequence
☑	(937) Putative DNA topoisomerase [Lactobacillus brevis]	3	MSIVVLTEKSSQIDPYAKALGTPKKKQGSVWVHSALL NADVQFVAASGHLFKLDGPEDYNDAWRDRSNYEQL PMVPQKFKARIKDARARASFNRIKQAITAADFVILATD
☑	(809) DNA topoisomerase 1 [Lactobacillus acidophilus NCFM]	1	MPTKAKSKSKTKKRKKILVIVESPAKAKTIEKYLGRNY HVIASKGHIRDLPKSQMGIDFEDNYKPKYISIRGKGDT IKELKSEAKKAKEVYLASDPDREGEAIAWHVAHALDL
☑	(276) Reverse gyrase [Caldicellulosiruptor saccharolyticus DSM 8903]	6	MATGAKYYHSCINCGGINSDTRNQKGLPCEKCLPFE NNMNILESLKEHNSLKDYAYFWNFQNEYKEFEKFFV SKVGKPMTGYQRLWARRLLLSKSFTLIAPTGLGKTTF

Showing 1 to 10 of 414 entries

Previous 1 2 3 4 5 … 42 Next

Family distribution



Taxonomy tree

- Firmicutes
  - Clostridia
  - Bacilli
  - Erysipelotrichia
  - Negativicutes

Figure: Taxon page.

Selections

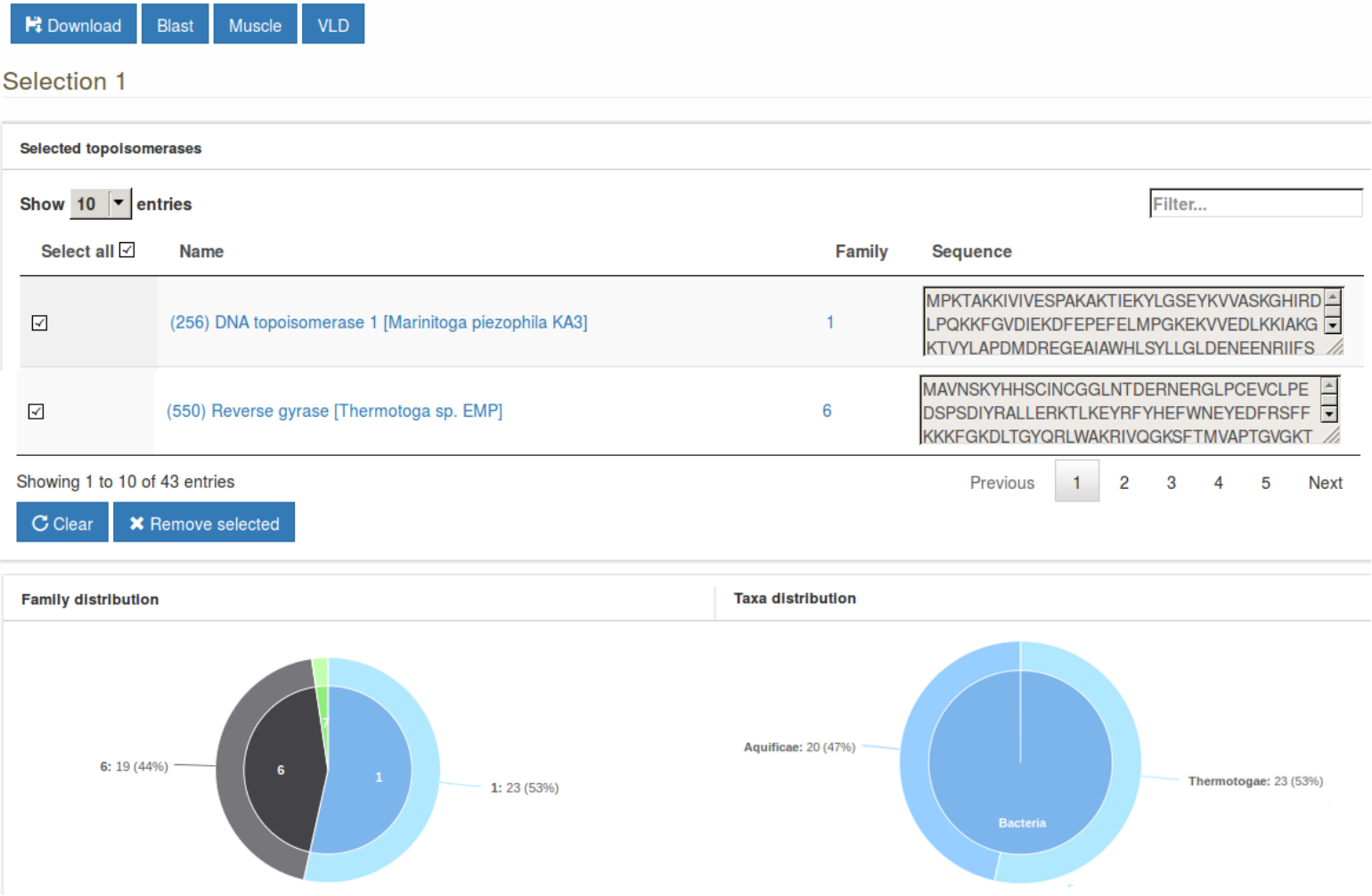


Figure: Selection page.

BLASTP

```
>90|DNA topoisomerase 1|1|AAA68949.1|P46799
MSKKVKKYIVVESPAKAKTIKSILGNEYEVFASMGHIIDLPKSKFGVDLEKDFEPEFAVI
KGKEKVVEKLKDLAKKGELLIASDMDREGEAIAWHIARVTNTLGRKNRIVFSEITPRVIR
EAVKNPREIDMKKVRAQLARRILDRIVGYSLSPVLWRNFKSNLSAGRVQSATLKLVCDRE
REILRFVPKKYHRITVNFDGLTAEIDVKEKKFFDAETLKEIQSIDELVVEEKKVSVKKFA
PPEPFKTSTLQQEAYSKLGFSVSKTMMIAQQLYEGVETKDGHIAFITYMRTDSTRVSDYA
KEEARNLITEVFGEEYVGSKRERRKSNAKIQDAHEAIRPTNVFMTPEEAGKYLNSDQKKL
YELIWKRFLASQMKPSQYEETRFVLRTKDGKYRFKGTVLKKIFDGYEKVWKTERNTGEFP
FEEGESVKPVVVKIEEQETKPKPRYTEGSLVKEMERLGIGRPSTYASTIKLLLNRGYIKK
IRGYLYPTIVGSVVMDYLEKKYSDVVSVSFTAEMEKDLDEVEQGKKTDKIVLREFYESFS
SVFDRNDRIVVDFPTNQKCSCGKEMRLSFGKYGFYLKCECGKTRSVKNDEIAVIDDGKIF
LGRKDSESGSPDGRSVEGKGNLSEKRRKGKKGS
>91|DNA topoisomerase 1|1|AAD35346.1|P46799
MSKKVKKYIVVESPAKAKTIKSILGNEYEVFASMGHIIDLPKSKFGVDLEKDFEPEFAVI
KGKEKVVEKLKDLAKKGELLIASDMDREGEAIAWHIARVTNTLGRKNRIVFSEITPRVIR
EAVKNPREIDMKKVRAQLARRILDRIVGYSLSPVLWRNFKSNLSAGRVQSATLKLVCDRE
REILRFVPKKYHRITVNFDGLTAEIDVKEKKFFDAETLKEIQSIDELVVEEKKVSVKKFA
PPEPFKTSTLQQEAYSKLGFSVSKTMMIAQQLYEGVETKDGHIAFITYMRTDSTRVSDYA
KEEARNLITEVFGEEYVGSKRERRKSNAKIQDAHEAIRPTNVFMTPEEAGKYLNSDQKKL
YELIWKRFLASQMKPSQYEETRFVLRTKDGKYRFKGTVLKKIFDGYEKVWKTERNTGEFP
FEEGESVKPVVVKIEEQETKPKPRYTEGSLVKEMERLGIGRPSTYASTIKLLLNRGYIKK
```

Load sample data

Parameters

Blast program	BLASTP
E-value	
Bank	selection1
Submit	Browse... No files selected.

Figure: BLASTP interface.

Other use-cases

- Taxa and VLD distribution of topoisomerases containing a specific amino acid sequence
- Analysis of a list of uniprot IDs of topoisomerases
- Classification of a new protein compared to stored topoisomerases

Perspectives

- Extension of the database to type IB and II topoisomerases
- Visualization of neighbors genomic elements
- Binding to other databases

Thank you for your attention
`stat.genopole.cnrs.fr/topodb`

Acknowledgements

Claudine Devauchelle and Franck Samson
Marc Nadal, Helène Debat, Eduardo Corel, Damien Correia,
Vladimir Daric, Anaïs Louis and Florence Vogliolo