

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

Nicolas DACCORD¹, Eduardo COREL^{1,2}, Damien CORREIA^{1,4}, Anais LOUIS¹, H el ene DEBAT^{3,4,5}, Vladimir DARIC⁴, Marc NADAL^{4,5}, Claudine DEVAUCHELLE¹ and Franck SAMSON¹

¹LaMME, UMR8071 CNRS, 23 bvd de France, Universit e d'Evry-Val d'Essonne |BGB|, 91037 Evry

²IBPS, UMR 7138, Universit e Pierre et Marie Curie, 7 quai Saint-Bernard, 75252 Paris

³Universit e Versailles St-Quentin, 45 avenue des Etats-Unis, 78035 Versailles

⁴IGM, UMR 8621 CNRS, Universit e Paris-Sud, B at. 409, 91405 Orsay Cedex

⁵Institut Jacques Monod, UMR 7592 CNRS, Universit e Paris Diderot



Topoisomerases

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

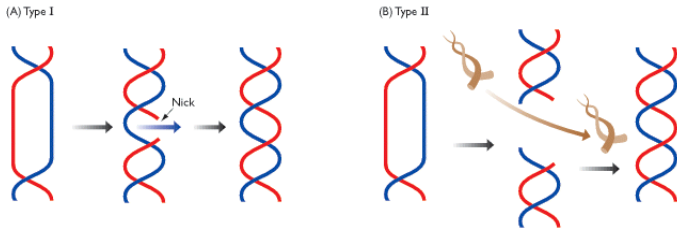


Figure: Mechanisms of action of type I et II topoisomerases.

Diversity

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

Topoisomerases			Characteristics			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	+	-	-	-
		TopoR	+	+	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

¹Didier, 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

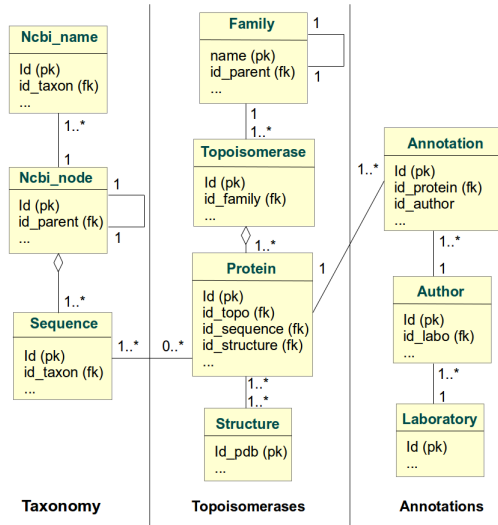


Figure: Simplified UML diagram of the database.

Data selection and expertise

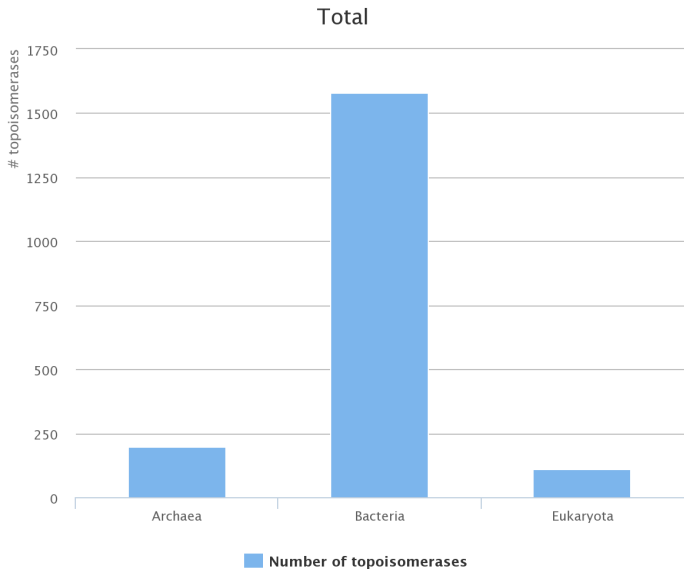
- Topoisomerases : TBLASTN on fully sequenced genomes
- Classification : VLD (Didier et al. 2012 ¹) + hierarchical ascendant clustering
- Annotations by experts

¹Didier, 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)

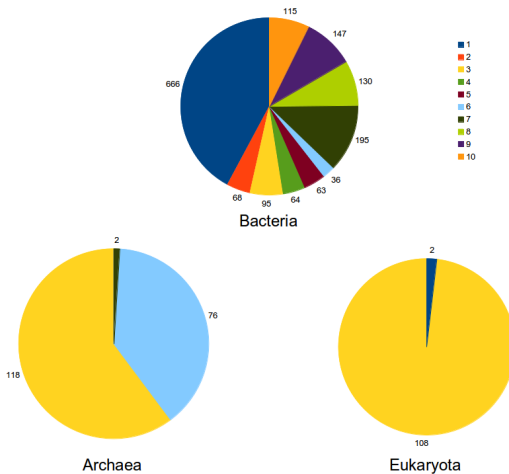


Figure: Topoisomerases distribution in domains.

Graphical interface

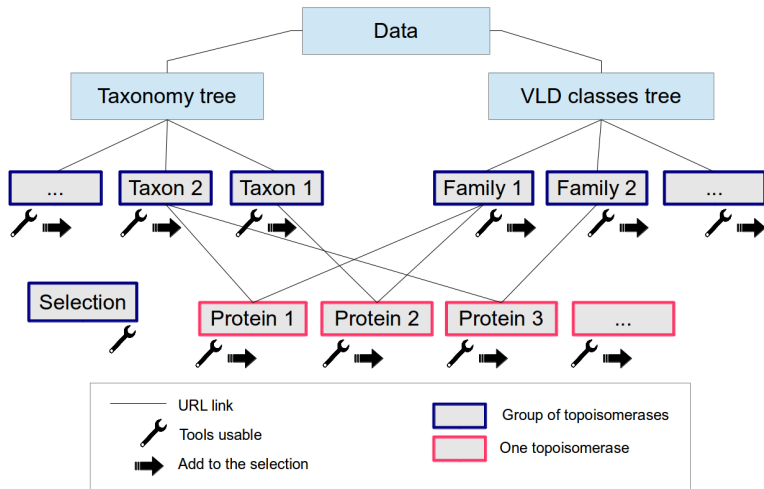


Figure: Data browsing on the interface.

Taxonomy

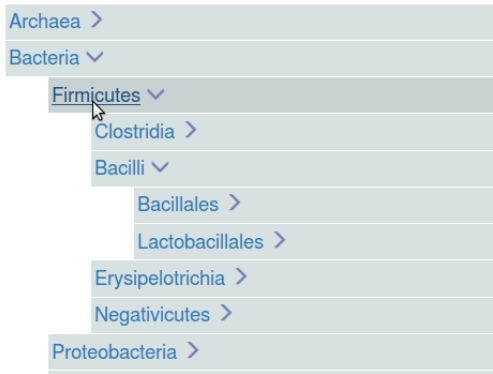


Figure: Part of the taxonomy tree.

Taxonomy

Download
Blast
Muscle
VLD
Add to selection ↗

Taxon : Firmicutes

Select all	Name	Family	Sequence
<input checked="" type="checkbox"/>	(937) Putative DNA topoisomerase [Lactobacillus brevis]	3	MSIVVLTEKSSQIDPYAKALGTPKKKQGSVWVHSALL NADVQFVAASGHLFKLDGPEYNDAWRDRSNYEQL PMVPOKFKARIKDARARASFNRRIKOAITAADEVII.ATD
<input checked="" type="checkbox"/>	(809) DNA topoisomerase 1 [Lactobacillus acidophilus NCFM]	1	MPTKAKSKSKTKKRRKILVIVESPAKAKTIEKYLGRNY HVIASKGHIRDLPKSQMGIDFEDNYKPKYISIRGKGD IKELKSEAKKAKEYVLASDPDREGEAIAWHVAHALDL
<input checked="" type="checkbox"/>	(276) Reverse gyrase [Caldicellulosiruptor saccharolyticus DSM 8903]	6	MATGAKYYYHSCINC GGINS DTRNQKGLPCEKLPFE NNMNI LESLKEHNSLKDYAYWNFQNEYKEFEKFFV SKVGKPM TGYQRLWARRLLLSKSF LIAPTGLGKTTF

Showing 1 to 10 of 414 entries

Previous 1 2 3 4 5 ... 42 Next

Family distribution

Family ID	Count	Percentage
1	206	50%
7	100	24%
2	68	16%
3	30	7%
6	10	2%

Taxonomy tree

- Firmicutes ▾
 - Clostridia >
 - Bacilli >
 - Erysipelotrichia >
 - Negativicutes >

Figure: Taxon page.

Selections

Download Blast Muscle VLD

Selection 1

Selected topoisomerases

Show **10** entries

Select all Name Family Sequence

<input checked="" type="checkbox"/>	(256) DNA topoisomerase 1 [Mariniflora piezophila KA3]	1	MPKTAKKIVIVESPAKAKTIEKYLGSEYKVVASKGHIRD LPQKKFGVDIEKDFEPEFELMPGKEKVVEDLKKIAG KTVYLLAPDMREGEIAIWHLSYLLGLDENEENRIIFS
<input checked="" type="checkbox"/>	(550) Reverse gyrase [Thermotoga sp. EMP]	6	MAVNSKYHHSCINCGLNTDERNERGLPCEVCLPE DSPSDIYRALLERKTLKEYRFYHEFWNEYEDFRSFF KKKFGKDLTGQYQRLWAKRIVQGKSFTMVAPTGVGKT

Showing 1 to 10 of 43 entries Previous **1** 2 3 4 5 Next

Family distribution

Family	Count	Percentage
1	23	53%
6	19	44%

Taxa distribution

Taxa	Count	Percentage
Bacteria	43	100%
Aquificae	20	47%
Thermotogae	23	53%


Figure: Selection page.


BLASTP


```
>98|DNA topoisomerase 1[1|AAA68949.1|P46799
MSKKVKKYIVVESPAKAKTIKSLGNEYEVFASMGHIIDLPKSKFGVDLEKDFEPEFAVI
KGKEKVVVEKLDLAKKGELLIASDMREGEAIWNIARVTNTLGRKNRIVFSEITPRVIR
EAVKNPREIDMKVRAQLARRILDRIVGYSLPVLWRNFKSNLSAGRVQSATLKLVCORE
REILRFVPKPYHRITVNFDDLTAEDVKEKKFFDAETLKEIQSIDELVVEKKVSVKKFA
PPEPFKSTLQQEAYSKLGFSVSKTMMIAQQLYEGVETKDGHIAFITYMRTDSTRVSDYA
KEEARNLITEVFGEEYVGSKRERRKSNAKIQDAHEAIRPTNVMFHTPEEAGYLNSDQKKL
YELIWKRFASQMKPSQYEETRFVLRKDGKGRFGTVLKKIFDGEKVMKTERNTGEFP
FEEGESVKKPVVVKIEEQETKPKPRYTEGSLVKEMERLIGRPSYASTIKLLNRGYIKK
IRGYLYPTIVGSMVDYLEKKYSDVSVSFTAEMEKDLDEVEQKKTKDKIVLREFYESFS
SVFDRNDRIVVDFPTNQKCSGKEMRLSFGKYGYLKCCEGKTRSVKNDIEAVIDDGKIF
LGRKDSSESGSDGRSVEGGNLSKRRKGGKGS
>91|DNA topoisomerase 1[1|AAD35346.1|P46799
MSKKVKKYIVVESPAKAKTIKSLGNEYEVFASMGHIIDLPKSKFGVDLEKDFEPEFAVI
KGKEKVVVEKLDLAKKGELLIASDMREGEAIWNIARVTNTLGRKNRIVFSEITPRVIR
EAVKNPREIDMKVRAQLARRILDRIVGYSLPVLWRNFKSNLSAGRVQSATLKLVCORE
REILRFVPKPYHRITVNFDDLTAEDVKEKKFFDAETLKEIQSIDELVVEKKVSVKKFA
PPEPFKSTLQQEAYSKLGFSVSKTMMIAQQLYEGVETKDGHIAFITYMRTDSTRVSDYA
KEEARNLITEVFGEEYVGSKRERRKSNAKIQDAHEAIRPTNVMFHTPEEAGYLNSDQKKL
YELIWKRFASQMKPSQYEETRFVLRKDGKGRFGTVLKKIFDGEKVMKTERNTGEFP
FEEGESVKKPVVVKIEEQETKPKPRYTEGSLVKEMERLIGRPSYASTIKLLNRGYIKK
```

[Load sample data](#)

Parameters

Blast program 

E-value 

Bank 

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, Hélène Debat, Eduardo Corel, Damien Correia,
Vladimir Daric, Anaïs Louis and Florence Vogliolo