



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.

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

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

¹LaMMe, UMR8071 CNRS, 23 bd de France, Université d'Evry-Val d'Essonne IBGBI, 91037 Evry

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

³Université Versailles St-Quentin, 45 avenue des Etats-Unis, 78035 Versailles

⁴IGM, UMR 8621 CNRS, Université Paris-Sud, Bât. 409, 91405 Orsay Cedex

⁵Institut Jacques Monod, UMR 7592 CNRS, Université Paris Diderot



Topoisomerases

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

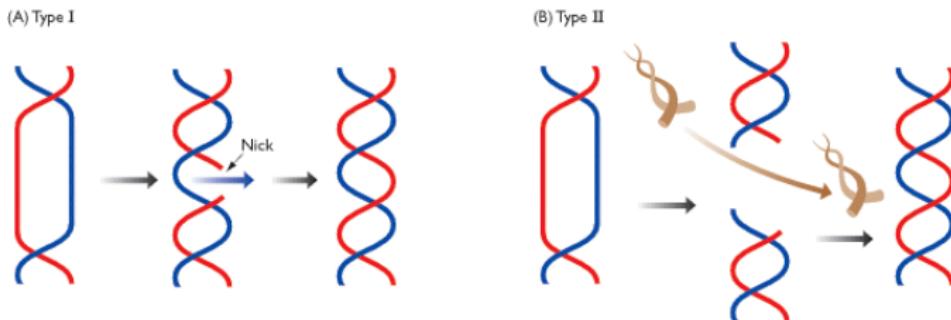


Figure: Mechanisms of action of type I et II topoisomerases.

Diversity

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

Topoisomerases			Characteristics			Activities			
Types	Sub-types	Name	Mg ²⁺	ATP	Cut	Releases negative supercoiling	Releases positive supercoiling	Inserts negative supercoiling	Inserts 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., 2012¹) 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

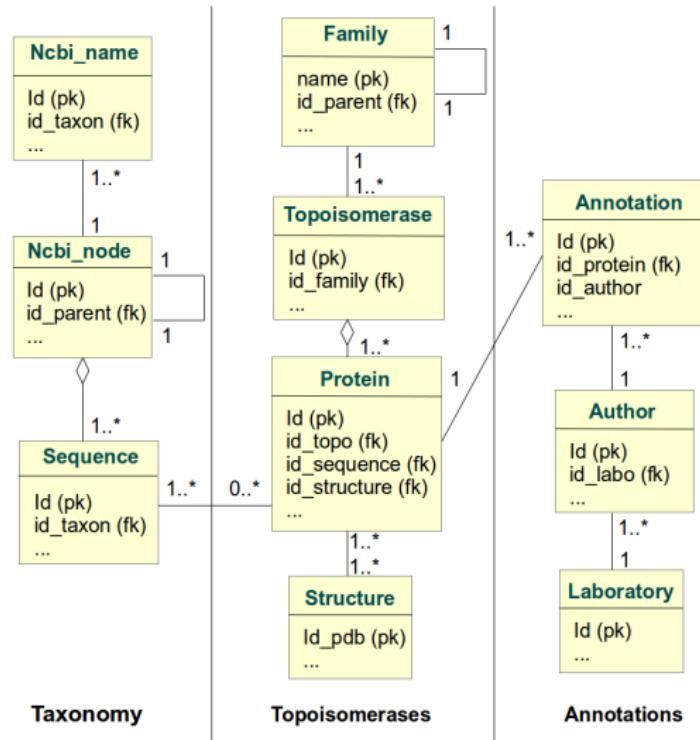


Figure: Simplified UML diagram of the database.

Data selection and expertise

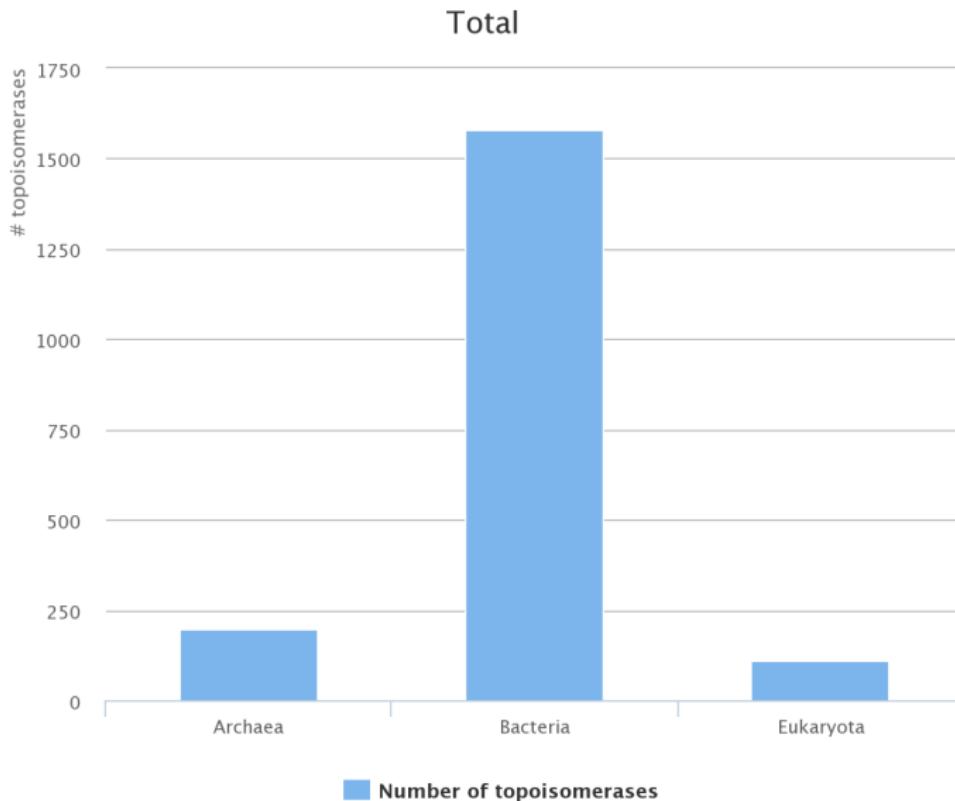
- Topoisomerases : TBLASTN on fully sequenced genomes
- Classification : VLD (Didier et al. 2012 ¹) + hierarchical ascendant clustering
- Annotations by experts

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)

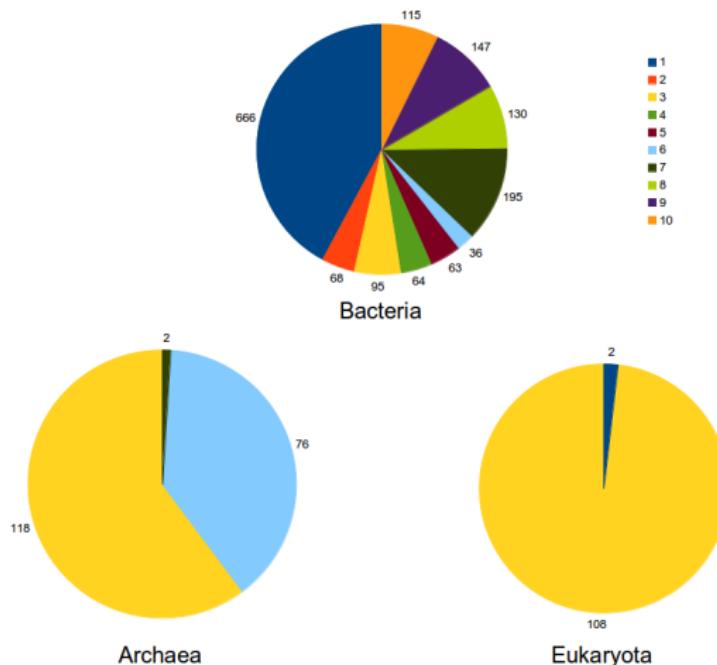


Figure: Topoisomerases distribution in domains.

Graphical interface

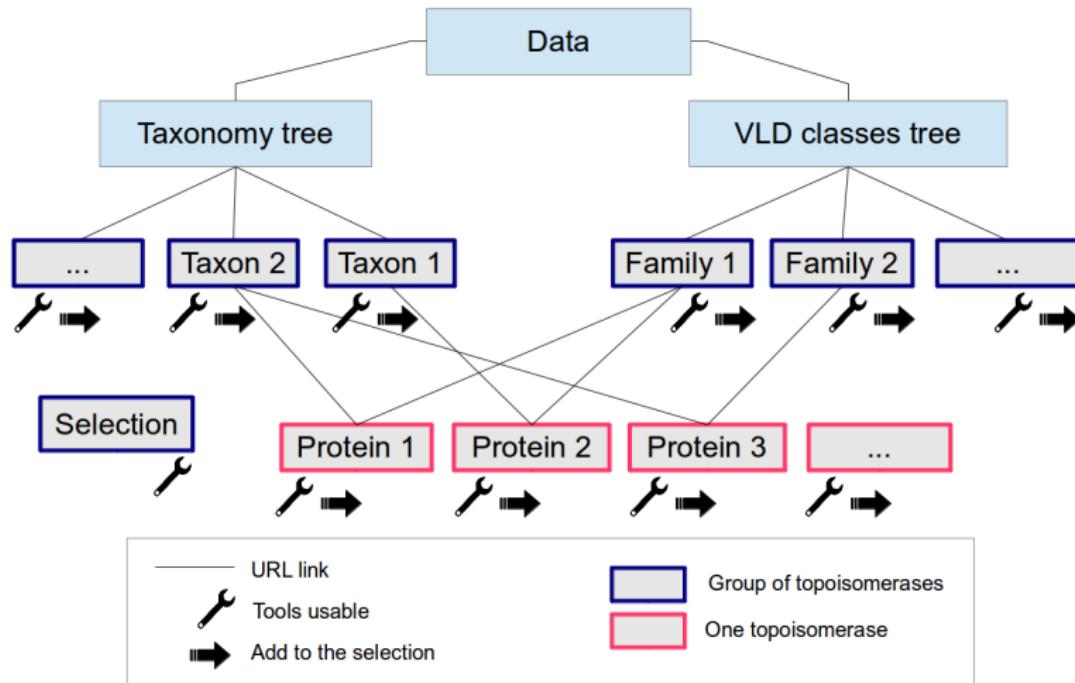


Figure: Data browsing on the interface.

Taxonomy

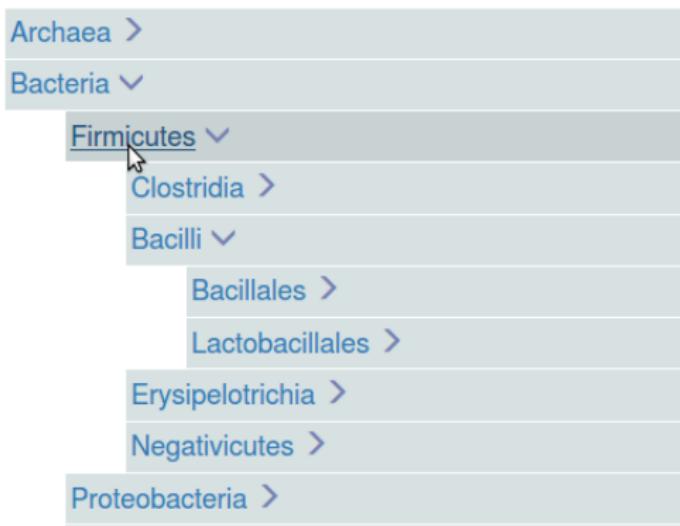


Figure: Part of the taxonomy tree.

Taxonomy

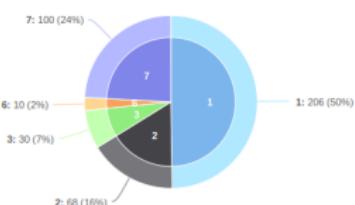
Taxon : Firmicutes

Select all	Name	Family	Sequence
<input checked="" type="checkbox"/>	(937) Putative DNA topoisomerase [Lactobacillus brevis]	3	MSIVVLTEKSSQIDPYAKALGTPKKQQGSVVHSALL NAD/QFVAASGHLFKLDPEDYNDAWRDRSNEYQL IPMVKPKFKARIKDARARASFNRNIKOAITAADFVIL ATD
<input checked="" type="checkbox"/>	(809) DNA topoisomerase 1 [Lactobacillus acidophilus NCFM]	1	MPTKAKSKSIKTKRKKILVIVESPAKAKTIEKYLGRNY HVIASKGHIRDLPLKSQMGIDFEDNYKPKYISIRGKGDT JKELKSEAKKAKEVYLASOPPDREGEIAWVVAHALDL
<input checked="" type="checkbox"/>	(276) Reverse gyrase [Caldicellulosiruptor saccharolyticus DSM 8903]	6	MATGAKYYHSCINC GGINS DTRNQKGLPCEKCLPFE NNMMIILLESKEHNSLKD YAYFWNFQNEYK EFEKFV ISKVGKPMGTQYQRLWARRLLLSSKSFTLIAPTGLKTF

Showing 1 to 10 of 414 entries

Previous 1 2 3 4 ... 42 Next

Family distribution



1: 206 (50%)
2: 68 (16%)
3: 30 (7%)
6: 10 (2%)
7: 100 (24%)

Taxonomy tree

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

Figure: Taxon page.

Selections

Download Blast Muscle VLD

Selection 1

Selected topoisomerases

Show 10 entries Filter...

	Name	Family	Sequence
<input checked="" type="checkbox"/>	(256) DNA topoisomerase 1 [Marinibacter piezophilus KA3]	1	MPTKAKKIVIVESPAKAKTIEKYLGSEYKVVASKGHIRD LPQKKFGVGDIEKDFEPEFELMPGKEKVVEDLKKIAKG KTVYLAPDMRDREGEAIAWHLSYLLGDLNEENRIIFS
<input checked="" type="checkbox"/>	(550) Reverse gyrase [Thermotoga sp. EMP]	6	MAVNSKYHHSCINCGLNTDERNERGLPCEVCLPE DSPSDIYRALLERKTLKEYFRFYHEFWNEYEDRSRFF KKKFGKDLTGQYQLWAKRIVQQKSFTMVAPTGVGKTT

Showing 1 to 10 of 43 entries

Previous 1 2 3 4 5 Next

Clear Remove selected

Family distribution Taxa distribution

1	23	53%
6	19	44%

Aquificae: 20	47%	
Bacteria	23	53%
Thermotogae: 23	53%	

Figure: Selection page.

BLASTP

```
>90|DNA topoisomerase 1|1|AAA68949.1|P46799
MSKKVKKYIVVESPAKAKTIKSILGNEYEVFASMGHIIDLPSKFGVDLEKDFEPEFAV
KGKEKVVKLKDOLAKKGELLIASMDREGEIAWHAIRVTNTLGRKNRIVFSEITPRVIR
EAVKNPREDMKKVRQLARRIIDRIVGYSLPVLWNRFNSNLAGRVSQSATLKLVCDR
REILRFPVKKYHRITVNFDGLTAEDIVKEKKFDAETLKEIQSIDEVLVEEKKVSVKFA
PPEPFKTSTLQQEAYSKLGFVSKTMMIAQQLYEGVETKDGHIAFITYMRDSTRVSDYA
KEEARNLITEVGEYVGSKRERRSNAKIQDAHEAIRPTNVPMTPEEAGKYLNSDQKLL
YELIWRFLASQMKPSQYEETRFLVLRKDGYRKFGTVLKKIFDGYEKVWKTERTNGEFP
FEEGESVKPVVKVIEEQETKPKPRVTEGSLVKEMERLGIGRPSTYASTIKLLLNRGYIKK
IRGYLYPTIVGVSVMDELYKKSDVSVSFTAEMLDDEVEQGKTDKIVLREFYESFS
SFDRNRDRIVVDFPTNQKCSGKEMRLSGFKYGFYLCECGKTRSVKNDEIAVIDDGKIF
LGRKDSESGSPGDRSVEKGKGMILSEKRKGKGS
>91|DNA topoisomerase 1|1|AAD35346.1|P46799
MSKKVKKYIVVESPAKAKTIKSILGNEYEVFASMGHIIDLPSKFGVDLEKDFEPEFAV
KGKEKVVKLKDOLAKKGELLIASMDREGEIAWHAIRVTNTLGRKNRIVFSEITPRVIR
EAVKNPREDMKKVRQLARRIIDRIVGYSLPVLWNRFNSNLAGRVSQSATLKLVCDR
REILRFPVKKYHRITVNFDGLTAEDIVKEKKFDAETLKEIQSIDEVLVEEKKVSVKFA
PPEPFKTSTLQQEAYSKLGFVSKTMMIAQQLYEGVETKDGHIAFITYMRDSTRVSDYA
KEEARNLITEVGEYVGSKRERRSNAKIQDAHEAIRPTNVPMTPEEAGKYLNSDQKLL
YELIWRFLASQMKPSQYEETRFLVLRKDGYRKFGTVLKKIFDGYEKVWKTERTNGEFP
FEEGESVKPVVKVIEEQETKPKPRVTEGSLVKEMERLGIGRPSTYASTIKLLLNRGYIKK
```

Load sample data

Parameters

Blast program	BLASTP
E-value	<input type="text"/>
Bank	selection1
<input type="button" value="Submit"/>	<input type="button" value="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, Hélène Debat, Eduardo Corel, Damien Correia,
Vladimir Daric, Anaïs Louis and Florence Vogliolo