

synthaser: a CD-Search enabled Python toolkit for analysing
domain architecture of fungal secondary metabolite
megasynt(et)ases

Cameron L.M. Gilchrist*¹, Yit-Heng Chooi*¹

¹School of Molecular Sciences, The University of Western Australia, 35 Stirling Hwy,
Crawley, 6009

E-mail: cameron.gilchrist@research.uwa.edu.au, yitheng.chooi@uwa.edu.au

List of Figures

S1	Schema of the synthaser JSON rule file.	2
S2	synthaser visualisation of domain architectures of non-reducing polyketide synthase (NR-PKS) sequences involved in meroterpenoid biosynthesis showing gaps in N-terminal regions of BAV69313.1 and EAU29529.1.	3
S3	Extract from multiple sequence alignment of N-terminal regions of non-reducing polyketide synthase (NR-PKS) sequences involved in meroterpenoid biosynthesis showing conservation of the starter unit:ACP transacylase (SAT) domain active site GXCXG motif.	4

```

{
  "domains": [
    "name": <string: name of the domain island>,
    "domains": [
      {
        "pssm": <integer: PSSM number>,
        "accession": <string: accession>,
        "name": <string: family name>,
        "length": <integer: PSSM length>,
        "bitscore": <float: threshold bitscore>,
        "superfamily": <string: superfamily>,
      }
    ]
  ],
  "rules": [
    {
      "name": <string: name of classification rule>,
      "domains": [<string: domain island name>],
      "filters": [
        {
          "type": <string: domain island name>,
          "domains": [<string: domain family accession to select>]
        }
      ],
      "renames": [
        {
          "from": <string: domain island being renamed>,
          "before": [<string: islands that must occur after renaming island>],
          "after": [<string: islands that must occur before renaming island>],
          "to": <string: new domain island name>
        }
      ],
      "evaluator": <string: rule evaluation string>,
    }
  ],
  "hierarchy": [
    {
      "title": <string: name of classification rule>,
      "children": [
        {
          "title": <string: name of child classification rule>,
          "children": [ ... ]
        }
      ]
    }
  ]
}

```

Figure S1: Schema of the `synthaser` JSON rule file. Fields used by the rule generator web application (unique identifiers, labels for HTML elements, etc) but not `synthaser` itself have been omitted for clarity. A more complete example is the fungal megasynthase rule set, which is distributed alongside the source code at <https://github.com/gamcil/synthaser> (`rules.json`).

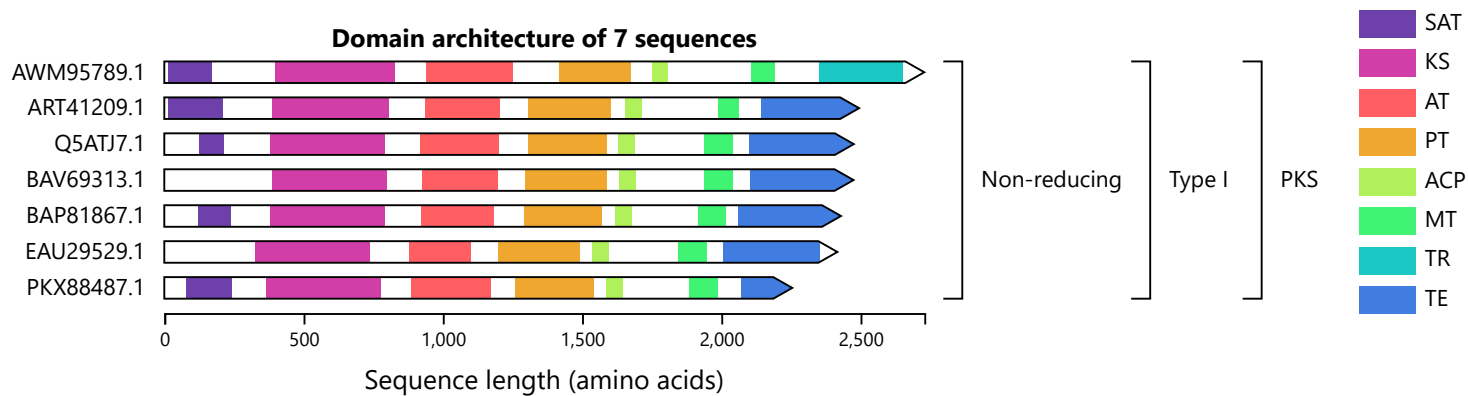


Figure S2: *synthaser* visualisation of domain architectures of non-reducing polyketide synthase (NR-PKS) sequences involved in meroterpenoid biosynthesis showing gaps in N-terminal regions of BAV69313.1 and EAU29529.1.

	120	130	140	150	160	170	180
AWM95789.1	RQSPTGLA	--EGKEAI	GFC TG	ILSAF	-AVASSHDVCDLAKY	GAAAMRLGMLVGLVVD	CED
Q5ATJ7.1	ELKEDK	-N-YDVCDI	GFC VG	FLAAIAAACWSDNE	DEFKVVSTVLR	LAVYIGAAVDL	DE
EAU29529.1	-----	-----	KKSE GFC VG	YLAAV	-AACWETDQTE	FPKAVATMLRI	AVCIGAVVDLDE
BAV69313.1	EFKEAGVN	-CDIKSM	GFC AG	YLAAV	-AACWEKDQSE	FSKVATMVRTAI	FIGAAVDLDE
ART41209.1	KFKEESQNR	CRIVDAQ	GFC VG	FLAAV	-AVASSNDSNE	FEDIAS	TMIRLAVCIGAAVDLDG
BAP81867.1	RLRKR	GSS-LQIKDV	GFC VG	FLAAT	-AVASAHDETQ	FRSIVAKVIR	LAVCVGALVDLNE
PKX88487.1	KVKETMKG	-FQARDV	GFC VG	FLAAT	-AVAASCD	ETAFRALVSKI	IRLAVCIGGLVDLDE

Figure S3: Extract from multiple sequence alignment of N-terminal regions of non-reducing polyketide synthase (NR-PKS) sequences involved in meroterpenoid biosynthesis showing conservation of the starter unit:ACP transacylase (SAT) domain active site GXCXG motif.