Supplementary Information for the exploration and data mining of fungal and oomycetes CAZomes

Emma E. M. Hobbs

June 2023

This document contains all supplementary information for chapter 6 in the thesis Hobbs, 2023. The tables and figures are presented in the same order as they are referenced in the main manuscript.

Contents

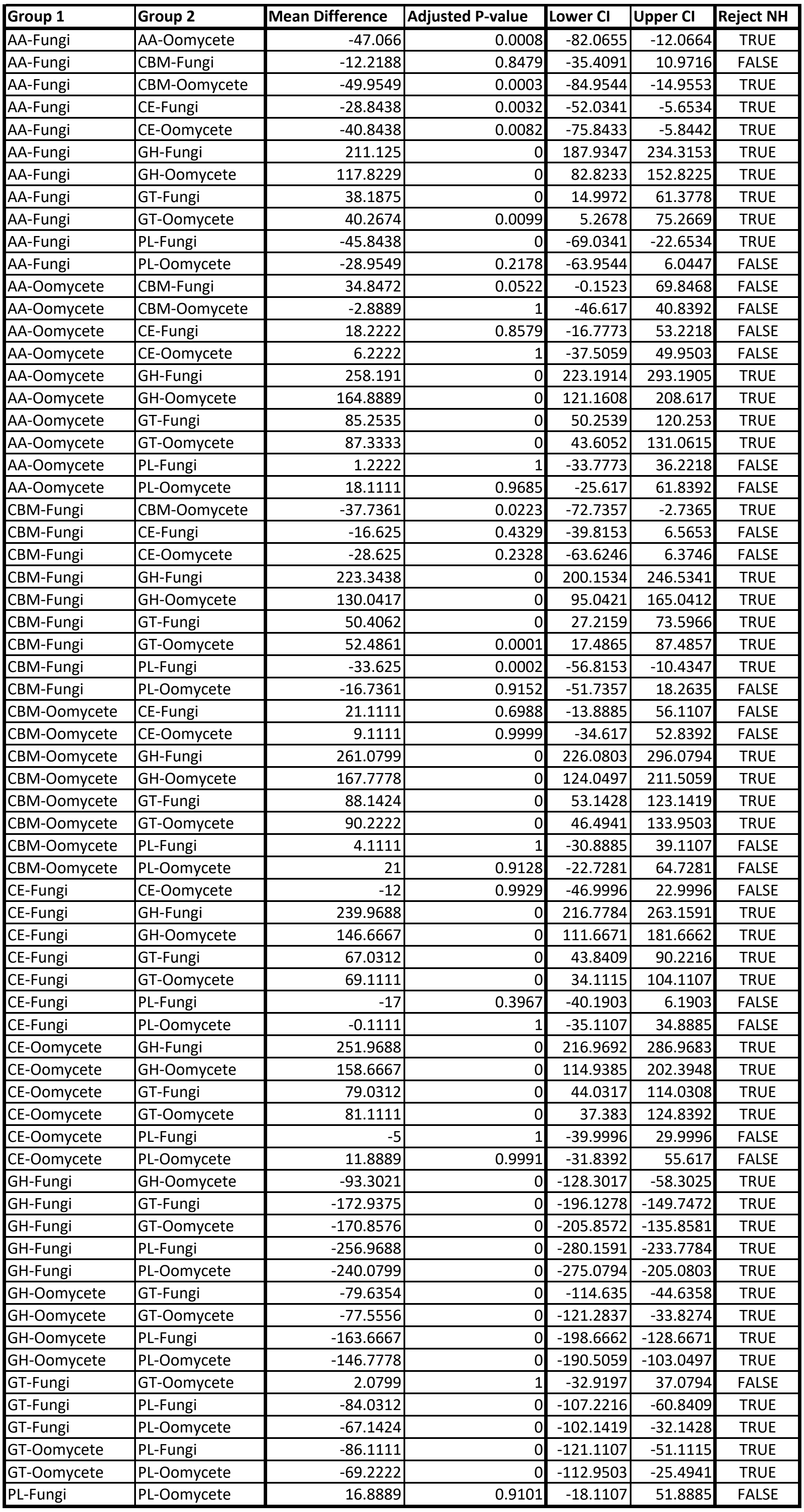
1 Genome download

SI Table 1: Genomic assemblies downloaded from NCBI

Genus	Species	NCBI Taxonomy ID	NCBI Accession Numbers
Aspergillus	*fumigatus*	NCBI:txid746128	GCA_012656185.1, GCA_012656215.1, GCA_012656165.1, GCA_012656115.1, GCA_012656125.1, GCA_005768625.2, GCA_003069565.1, GCA_002234985.1, GCA_002234955.1, GCA_001715275.2, GCA_001643655.1, GCA_001643665.1
Aspergillus	*nidulans*	NCBI:txid162425	GCA_011075025.1, GCA_011074995.1
Aspergillus	*niger*	NCBI:txid5061	GCA_011316255.1, GCA_009812365.1, GCA_004634315.1, GCA_002211485.2, GCA_900248155.1, GCA_002740505.1, GCA_001931795.1, GCA_001741915.1, GCA_001741905.1, GCA_001741885.1, GCA_001715265.1, GCA_001515345.1, GCF_000002855.3
Aspergillus	*sydowii*	NCBI:txid75750	GCA_009828905.1, GCA_009193685.1
Fusarium	*graminearum*	NCBI:txid5518	GCA_012959185.1, GCA_006942295.1, GCA_900492705.1, GCA_900476405.1, GCA_002352725.1, GCA_900044135.1, GCA_001717915.1, GCA_001717905.1, GCA_000966635.1, GCA_000966645.1, GCA_000599445.1
Fusarium	*oxysporum*	NCBI:txid5507	GCA_011428085.1, GCA_011426355.1, GCA_011426335.1, GCA_011424645.1, GCA_011424625.1, GCA_011424605.1, GCA_011421335.1, GCA_011421285.1, GCA_011421305.1, GCA_011421375.1, GCA_011421365.1, GCA_011421355.1, GCA_011421275.1, GCA_011421325.1, GCA_011037735.1, GCA_011037105.1, GCA_011037075.1, GCA_011036425.1, GCA_011036365.1, GCA_011036345.1, GCA_011036325.1, GCA_011036305.1, GCA_011036285.1, GCA_011036015.1, GCA_011035995.1, GCA_011035975.1, GCA_011035895.1, GCA_011035875.1, GCA_011035855.1, GCA_011035785.1, GCA_011035765.1, GCA_011035725.1, GCA_011037135.1, GCA_011037005.1, GCA_011036985.1, GCA_011036965.1, GCA_011035695.1, GCA_011036925.1, GCA_011036905.1, GCA_011036835.1, GCA_011035665.1, GCA_011035645.1, GCA_011035625.1, GCA_011035595.1, GCA_011035555.1, GCA_011035525.1, GCA_011036745.1, GCA_011035505.1, GCA_011036685.1, GCA_011036655.1, GCA_011036635.1, GCA_011036615.1, GCA_011035355.1, GCA_011036575.1, GCA_011035205.1, GCA_011035185.1, GCA_011035135.1, GCA_011035015.1, GCA_011034965.1, GCA_011034945.1, GCA_011034875.1, GCA_011034825.1, GCA_011034785.1, GCA_011034745.1, GCA_011034655.1, GCA_011034575.1, GCA_011034545.1, GCA_011034455.1, GCA_011034415.1, GCA_011034375.1, GCA_011034275.1, GCA_011034205.1, GCA_011034135.1, GCA_011034075.1, GCA_011034045.1, GCA_011034025.1, GCA_011037795.1, GCA_011033995.1, GCA_011033925.1, GCA_011033815.1, GCA_011033715.1, GCA_011033745.1, GCA_011033645.1, GCA_011036945.1, GCA_011036875.1, GCA_011036795.1, GCA_011036775.1, GCA_011036815.1, GCA_011036855.1, GCA_011036595.1, GCA_011036705.1, GCA_011036765.1, GCA_011036565.1, GCA_011036545.1, GCA_011036445.1, GCA_011036725.1, GCA_011036505.1, GCA_011036515.1, GCA_011036475.1, GCA_011036455.1, GCA_011036395.1, GCA_011036385.1, GCA_011036275.1, GCA_011036235.1, GCA_011036165.1, GCA_011036225.1, GCA_011036215.1, GCA_011036205.1, GCA_011036075.1, GCA_011036135.1, GCA_011036125.1, GCA_011036115.1, GCA_011036055.1, GCA_011036065.1, GCA_011036045.1, GCA_011035965.1, GCA_011035955.1, GCA_011035835.1, GCA_011035845.1, GCA_011035825.1, GCA_011035755.1, GCA_011035745.1, GCA_011033685.1, GCA_011033665.1, GCA_011033625.1, GCA_011033575.1, GCA_011033555.1, GCA_011033535.1, GCA_011033505.1, GCA_011033485.1, GCA_011033455.1, GCA_011035615.1, GCA_011035485.1, GCA_011035495.1, GCA_011035455.1, GCA_011035415.1, GCA_011035435.1, GCA_011035375.1, GCA_011035345.1, GCA_011035385.1, GCA_011035335.1, GCA_011035235.1, GCA_011035255.1, GCA_011035245.1, GCA_011035265.1, GCA_011035275.1, GCA_011035075.1, GCA_011035065.1, GCA_011035045.1, GCA_011035055.1, GCA_011035035.1, GCA_011034915.1, GCA_011034925.1, GCA_011034935.1, GCA_011034815.1, GCA_011034845.1, GCA_011034805.1, GCA_011034775.1, GCA_011034735.1, GCA_011034615.1, GCA_011034635.1, GCA_011034645.1, GCA_011034625.1, GCA_011034675.1, GCA_011034565.1, GCA_011034515.1, GCA_011034445.1, GCA_011034485.1, GCA_011034475.1, GCA_011034395.1, GCA_011034265.1, GCA_011034195.1, GCA_011034235.1, GCA_011034155.1, GCA_011034225.1, GCA_011033985.1, GCA_011033945.1, GCA_011034125.1, GCA_011034105.1, GCA_011033955.1, GCA_011033875.1, GCA_011034095.1, GCA_011033805.1, GCA_011033885.1, GCA_011033895.1, GCA_011033835.1, GCA_011033705.1, GCA_011033765.1, GCA_011033785.1, GCA_011033475.1, GCA_011033595.1, GCA_011033375.1, GCA_011033525.1, GCA_011033385.1, GCA_011032885.1, GCA_011032855.1, GCA_009746015.1, GCA_009299335.1, GCA_009299235.1, GCA_009299215.1, GCA_009299195.1, GCA_009299155.1, GCA_009299095.1, GCA_009299045.1, GCA_009298875.1, GCA_009298855.1, GCA_009298805.1, GCA_009298685.1, GCA_009298645.1, GCA_009298615.1, GCA_009298555.1, GCA_009298505.1, GCA_009298475.1, GCA_009298435.1, GCA_009298405.1, GCA_009298245.1, GCA_009298235.1, GCA_009298205.1, GCA_009298195.1, GCA_009298175.1, GCA_009298145.1, GCA_009298125.1, GCA_009298085.1, GCA_009298065.1, GCA_009298075.1, GCA_009298035.1, GCA_009297995.1, GCA_009297985.1, GCA_009297935.1, GCA_009297945.1, GCA_009297925.1, GCA_009297855.1, GCA_009297755.1, GCA_009297735.1, GCA_009297675.1, GCA_009297655.1, GCA_009297635.1, GCA_009297575.1, GCA_009297555.1, GCA_009297465.1, GCA_009297425.1, GCA_009297405.1, GCA_009297385.1, GCA_009297365.1, GCA_009297885.1, GCA_009297835.1, GCA_009299255.1, GCA_009299295.1, GCA_009299135.1, GCA_009299115.1, GCA_009299075.1, GCA_009299125.1, GCA_009299175.1, GCA_009298955.1, GCA_009299025.1, GCA_009298985.1, GCA_009299005.1, GCA_009298915.1, GCA_009298925.1, GCA_009298935.1, GCA_009298945.1, GCA_009298845.1, GCA_009298825.1, GCA_009298675.1, GCA_009298715.1, GCA_009298755.1, GCA_009298705.1, GCA_009298745.1, GCA_009298655.1, GCA_009298635.1, GCA_009298515.1, GCA_009298545.1, GCA_009298495.1, GCA_009298455.1, GCA_009298465.1, GCA_009298395.1, GCA_009298275.1, GCA_009298295.1, GCA_009298315.1, GCA_009298335.1, GCA_009298285.1, GCA_009298305.1, GCA_009298045.1, GCA_009297915.1, GCA_009297825.1, GCA_009297725.1, GCA_009297785.1, GCA_009297715.1, GCA_009297695.1, GCA_009297625.1, GCA_009297605.1, GCA_009297515.1, GCA_009297505.1, GCA_009297445.1, GCA_009297485.1, GCA_009297495.1, GCA_004292535.1, GCA_004291455.1, GCA_003709395.1, GCA_003705045.1, GCA_003704975.1, GCA_003705035.1, GCA_003615165.1, GCA_003615155.1, GCA_003615115.1, GCA_003615185.1, GCA_003025235.1, GCA_003025205.1, GCA_002894245.1, GCA_900096695.1, GCA_002233955.1, GCA_002233985.1, GCA_002233935.1, GCA_002233995.1, GCA_001931975.2, GCA_001703125.1, GCA_000733055.2
Fusarium	*proliferatum*	NCBI:txid948311	GCA_003709405.1, GCA_003705095.1, GCA_003704965.1, GCA_003704895.1, GCA_003704885.1, GCA_003704875.1, GCA_003615215.1, GCA_003290285.1, GCA_003123625.1, GCA_002234285.1, GCA_900029915.1, GCA_001705295.1
Magnaporthe	*grisea*	NCBI:txid148305	GCF_004355905.1, GCA_003933175.1, GCA_002925245.1, GCA_002924675.1, GCA_001548815.1, GCA_001548795.1
Magnaporthe	*oryzae*	NCBI:txid318829	GCA_012979135.1, GCA_012978465.1, GCA_012978415.1, GCA_012979075.1, GCA_012978505.1, GCA_012978515.1, GCA_012978495.1, GCA_012978435.1, GCA_012272995.1, GCA_012922935.1, GCA_012654135.1, GCA_012654105.1, GCA_012654075.1, GCA_012654115.1, GCA_012654035.1, GCA_012596185.1, GCA_012490815.1, GCA_012490805.1, GCA_011799965.1, GCA_011799925.1, GCA_011799915.1, GCA_011799905.1, GCA_900474545.3, GCA_900474475.3, GCA_900474655.3, GCA_900474175.3, GCA_004785725.1, GCA_004346965.1, GCA_900474375.2, GCA_900474635.2, GCA_900474435.2, GCA_900474225.2, GCA_003991345.1, GCA_003017255.1, GCA_003017175.1, GCA_003017165.1, GCA_003017125.1, GCA_003017115.1, GCA_003017045.1, GCA_003017065.1, GCA_003017035.1, GCA_003017025.1, GCA_003016985.1, GCA_003016965.1, GCA_003016955.1, GCA_003016935.1, GCA_003016905.1, GCA_003016895.1, GCA_003016875.1, GCA_003016855.1, GCA_003016825.1, GCA_003016805.1, GCA_003016795.1, GCA_003016785.1, GCA_003016745.1, GCA_003016725.1, GCA_003016715.1, GCA_003016705.1, GCA_003016665.1, GCA_003016655.1, GCA_003016635.1, GCA_003016625.1, GCA_003016585.1, GCA_003016555.1, GCA_003016575.1, GCA_003016545.1, GCA_003016505.1, GCA_003016495.1, GCA_003016475.1, GCA_003016465.1, GCA_003016425.1, GCA_003016415.1, GCA_003016395.1, GCA_003016385.1, GCA_003016325.1, GCA_003016265.1, GCA_003016275.1, GCA_003016255.1, GCA_003016245.1, GCA_003016195.1, GCA_003016185.1, GCA_003016175.1, GCA_003016165.1, GCA_003016105.1, GCA_003016115.1, GCA_003016095.1, GCA_003016085.1, GCA_003016015.1, GCA_003016035.1, GCA_003016025.1, GCA_003016005.1, GCA_003015975.1, GCA_003015955.1, GCA_003015935.1, GCA_003015925.1, GCA_003015895.1, GCA_003015885.1, GCA_003015825.1, GCA_003015835.1, GCA_003015815.1, GCA_003015805.1, GCA_003015755.1, GCA_003015745.1, GCA_003015735.1, GCA_003015705.1, GCA_003015645.1, GCA_003015655.1, GCA_003015635.1, GCA_003015625.1, GCA_003015595.1, GCA_003015565.1, GCA_003015555.1, GCA_003015545.1, GCA_003015495.1, GCA_003015515.1, GCA_003015465.1, GCA_003015475.1, GCA_003015425.1, GCA_003015405.1, GCA_003015395.1, GCA_003015385.1, GCA_003013125.1, GCA_002924695.1, GCA_002925445.1, GCA_002925415.1, GCA_002925425.1, GCA_002925405.1, GCA_002925385.1, GCA_002925325.1, GCA_002925335.1, GCA_002925345.1, GCA_002925295.1, GCA_002925285.1, GCA_002925215.1, GCA_002925225.1, GCA_002925205.1, GCA_002925165.1, GCA_002925145.1, GCA_002925155.1, GCA_002925095.1, GCA_002925085.1, GCA_002925105.1, GCA_002925065.1, GCA_002925045.1, GCA_002924965.1, GCA_002925025.1, GCA_002924985.1, GCA_002924975.1, GCA_002924945.1, GCA_002924885.1, GCA_002924865.1, GCA_002924915.1, GCA_002924875.1, GCA_002924825.1, GCA_002924785.1, GCA_002924835.1, GCA_002924795.1, GCA_002924755.1, GCA_002924745.1, GCA_002924665.1, GCA_002924685.1, GCA_002368515.1, GCA_002368525.1, GCA_002368485.1, GCA_002368475.1, GCA_002218485.1, GCA_002218465.1, GCA_002218475.1, GCA_002218435.1, GCA_002218425.1, GCA_002218355.1, GCA_002218345.1, GCA_002105295.1, GCA_001936935.1, GCA_001936435.1, GCA_001936075.1, GCA_001853415.2, GCA_001675605.1, GCA_001675625.1, GCA_001675595.1, GCA_001675615.1, GCA_001548855.1, GCA_001548845.1, GCA_001548775.1, GCA_001548785.1, GCA_000805855.1, GCA_000734785.1, GCA_000734755.1, GCA_000734735.1, GCA_000734685.1, GCA_000734675.1, GCA_000734705.1, GCA_000734655.1, GCA_000734635.1, GCA_000734605.1, GCA_000734595.1, GCA_000734575.1, GCA_000734555.1, GCA_000734515.1, GCA_000734525.1, GCA_000734495.1, GCA_000734455.1, GCA_000734425.1, GCA_000734395.1, GCA_000734405.1, GCA_000734325.1, GCA_000734345.1, GCA_000734335.1, GCA_000734315.1, GCA_000734275.1, GCA_000734265.1, GCA_000734245.1, GCA_000734235.1, GCA_000734215.1, GCA_000734185.1, GCA_000734165.1, GCA_000734155.1, GCA_000734105.1, GCA_000734075.1, GCA_000734095.1, GCA_000734085.1
Mycosphaerella	*graminicola*	NCBI:txid1047171	GCA_902712725.1, GCA_003613095.1, GCA_003611185.1, GCA_003611175.1, GCA_003611135.1, GCA_003611115.1, GCA_003611125.1, GCA_003611075.1, GCA_003611065.1, GCA_003611055.1, GCA_002937425.1
Rhynchosporium	*agropyri*	NCBI:txid914238	GCA_900074905.1
Rhynchosporium	*commune*	NCBI:txid914237	GCA_900074885.1
Rhynchosporium	*secalis*	NCBI:txid38038	GCA_900074895.1
Trichoderma	*asperellum*	NCBI:txid101201	GCA_004154885.1, GCA_000733085.2
Trichoderma	*atroviride*	NCBI:txid63577	GCA_002916895.1, GCA_001599035.1, GCA_000963795.1
Trichoderma	*citrinoviride*	NCBI:txid58853	GCF_003025115.1
Trichoderma	*harzianum*	NCBI:txid5544	GCA_010015525.1, GCA_002894145.1, GCA_002838845.1, GCA_001990665.1, GCA_000988865.1
Trichoderma	*reesei*	NCBI:txid51453	GCA_004762065.1, GCA_001999515.1
Ustilago	*maydis*	NCBI:txid5270	GCA_001736185.1, GCA_001736215.1, GCA_001736155.1, GCA_001662005.1, GCA_001660065.1, GCA_001599495.1
Ustilago	*bromivora*	NCBI:txid307758	GCA_900101485.1, GCA_900080155.1
Albugo	*candida*	NCBI:txid65357	GCA_000326065.1, GCA_000326045.1, GCA_001306775.1, GCA_001306755.1, GCA_000313105.1, GCA_001078535.1, GCA_000961115.1
Hyaloperonospora	*arabidopsidis*	NCBI:txid272952	GCA_001414525.1, GCA_001414265.1
Phytophthora	*cinnamomi*	NCBI:txid4785	GCA_002734105.1, GCA_002734125.1, GCA_001314505.1, GCA_001314365.1
Phytophthora	*capsici*	NCBI:txid4784	GCA_004138045.1, GCA_004137965.1, GCA_004137975.1, GCA_004137955.1, GCA_004137885.1, GCA_004137865.1
Phytophthora	*infestans*	NCBI:txid4787	GCA_012552325.1, GCA_012295175.1, GCA_011316315.1, GCA_001661535.1
Phytophthora	*parasitica*	NCBI:txid4792	GCA_000509525.1, GCA_000509505.1, GCA_000509465.1, GCA_000509485.1
Phytophthora	*ramorum*	NCBI:txid164328	GCA_004343245.1, GCA_003956735.1, GCA_002968915.1, GCA_000340395.2, GCA_001955675.1, GCA_001933465.1, GCA_001933485.1, GCA_001933455.1, GCA_001933415.1, GCA_001933395.1, GCA_001933405.1, GCA_001933345.1, GCA_001933325.1, GCA_001933315.1, GCA_001933335.1, GCA_000336535.2, GCA_001278225.1, GCA_001278215.1, GCA_001278235.1, GCA_001278165.1, GCA_001278155.1, GCA_001278135.1, GCA_001278145.1, GCA_000149735.1
Phytophthora	*sojae*	NCBI:txid67593	GCA_009848525.1, GCF_000149755.1
Plasmopara	*viticola*	NCBI:txid143451	GCA_001695595.3, GCA_003123765.1, GCA_001974925.1
Plasmopara	*halstedii*	NCBI:txid4781	GCA_004380875.1, GCA_003724065.1, GCA_003640465.1, GCA_003640505.1, GCF_900000015.1
Plasmopara	*obducens*	NCBI:txid162140	GCA_003640625.1, GCA_003640485.1

2 Exploration of the fungi and oomycetes CAZomes

SI Table 2: Output of Tukey HSD test

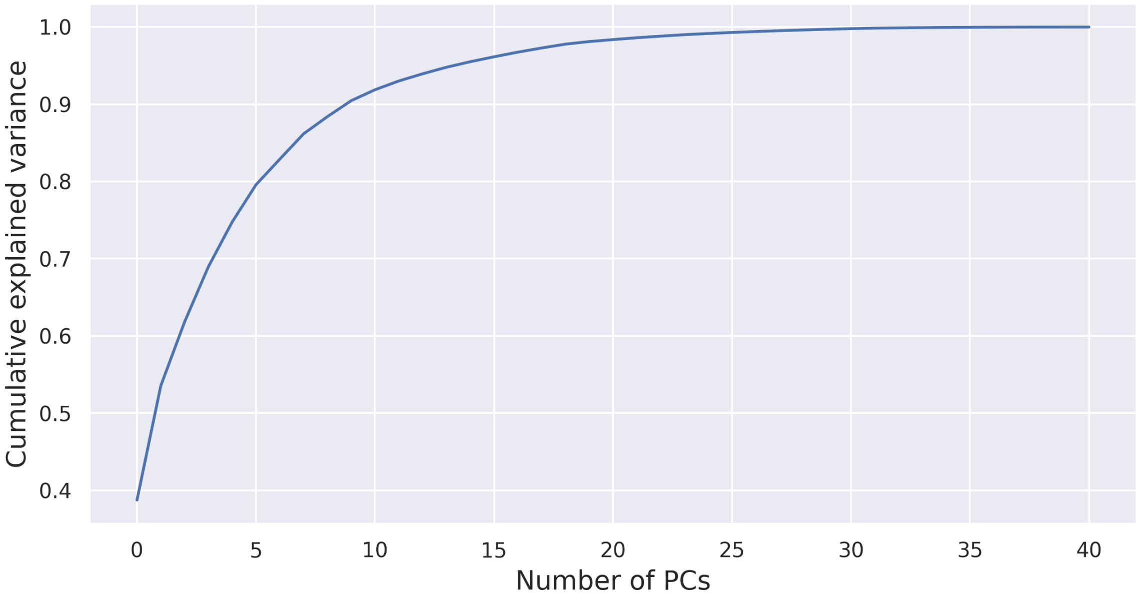


Group 1	Group 2	Mean Difference	Adjusted P-value	Lower CI	Upper CI	Reject NH
AA-Fungi	AA-Oomycete	-47.066	0.0008	-82.0655	-12.0664	TRUE
AA-Fungi	CBM-Fungi	-12.2188	0.8479	-35.4091	10.9716	FALSE
AA-Fungi	CBM-Oomycete	-49.9549	0.0003	-84.9544	-14.9553	TRUE
AA-Fungi	CE-Fungi	-28.8438	0.0032	-52.0341	-5.6534	TRUE
AA-Fungi	CE-Oomycete	-40.8438	0.0082	-75.8433	-5.8442	TRUE
AA-Fungi	GH-Fungi	211.125	0	187.9347	234.3153	TRUE
AA-Fungi	GH-Oomycete	117.8229	0	82.8233	152.8225	TRUE
AA-Fungi	GT-Fungi	38.1875	0	14.9972	61.3778	TRUE
AA-Fungi	GT-Oomycete	40.2674	0.0099	5.2678	75.2669	TRUE
AA-Fungi	PL-Fungi	-45.8438	0	-69.0341	-22.6534	TRUE
AA-Fungi	PL-Oomycete	-28.9549	0.2178	-63.9544	6.0447	FALSE
AA-Oomycete	CBM-Fungi	34.8472	0.0522	-0.1523	69.8468	FALSE
AA-Oomycete	CBM-Oomycete	-2.8889	1	-46.617	40.8392	FALSE
AA-Oomycete	CE-Fungi	18.2222	0.8579	-16.7773	53.2218	FALSE
AA-Oomycete	CE-Oomycete	6.2222	1	-37.5059	49.9503	FALSE
AA-Oomycete	GH-Fungi	258.191	0	223.1914	293.1905	TRUE
AA-Oomycete	GH-Oomycete	164.8889	0	121.1608	208.617	TRUE
AA-Oomycete	GT-Fungi	85.2535	0	50.2539	120.253	TRUE
AA-Oomycete	GT-Oomycete	87.3333	0	43.6052	131.0615	TRUE
AA-Oomycete	PL-Fungi	1.2222	1	-33.7773	36.2218	FALSE
AA-Oomycete	PL-Oomycete	18.1111	0.9685	-25.617	61.8392	FALSE
CBM-Fungi	CBM-Oomycete	-37.7361	0.0223	-72.7357	-2.7365	TRUE
CBM-Fungi	CE-Fungi	-16.625	0.4329	-39.8153	6.5653	FALSE
CBM-Fungi	CE-Oomycete	-28.625	0.2328	-63.6246	6.3746	FALSE
CBM-Fungi	GH-Fungi	223.3438	0	200.1534	246.5341	TRUE
CBM-Fungi	GH-Oomycete	130.0417	0	95.0421	165.0412	TRUE
CBM-Fungi	GT-Fungi	50.4062	0	27.2159	73.5966	TRUE
CBM-Fungi	GT-Oomycete	52.4861	0.0001	17.4865	87.4857	TRUE
CBM-Fungi	PL-Fungi	-33.625	0.0002	-56.8153	-10.4347	TRUE
CBM-Fungi	PL-Oomycete	-16.7361	0.9152	-51.7357	18.2635	FALSE
CBM-Oomycete	CE-Fungi	21.1111	0.6988	-13.8885	56.1107	FALSE
CBM-Oomycete	CE-Oomycete	9.1111	0.9999	-34.617	52.8392	FALSE
CBM-Oomycete	GH-Fungi	261.0799	0	226.0803	296.0794	TRUE
CBM-Oomycete	GH-Oomycete	167.7778	0	124.0497	211.5059	TRUE
CBM-Oomycete	GT-Fungi	88.1424	0	53.1428	123.1419	TRUE
CBM-Oomycete	GT-Oomycete	90.2222	0	46.4941	133.9503	TRUE
CBM-Oomycete	PL-Fungi	4.1111	1	-30.8885	39.1107	FALSE
CBM-Oomycete	PL-Oomycete	21	0.9128	-22.7281	64.7281	FALSE
CE-Fungi	CE-Oomycete	-12	0.9929	-46.9996	22.9996	FALSE
CE-Fungi	GH-Fungi	239.9688	0	216.7784	263.1591	TRUE
CE-Fungi	GH-Oomycete	146.6667	0	111.6671	181.6662	TRUE
CE-Fungi	GT-Fungi	67.0312	0	43.8409	90.2216	TRUE
CE-Fungi	GT-Oomycete	69.1111	0	34.1115	104.1107	TRUE
CE-Fungi	PL-Fungi	-17	0.3967	-40.1903	6.1903	FALSE
CE-Fungi	PL-Oomycete	-0.1111	1	-35.1107	34.8885	FALSE
CE-Oomycete	GH-Fungi	251.9688	0	216.9692	286.9683	TRUE
CE-Oomycete	GH-Oomycete	158.6667	0	114.9385	202.3948	TRUE
CE-Oomycete	GT-Fungi	79.0312	0	44.0317	114.0308	TRUE
CE-Oomycete	GT-Oomycete	81.1111	0	37.383	124.8392	TRUE
CE-Oomycete	PL-Fungi	-5	1	-39.9996	29.9996	FALSE
CE-Oomycete	PL-Oomycete	11.8889	0.9991	-31.8392	55.617	FALSE
GH-Fungi	GH-Oomycete	-93.3021	0	-128.3017	-58.3025	TRUE
GH-Fungi	GT-Fungi	-172.9375	0	-196.1278	-149.7472	TRUE
GH-Fungi	GT-Oomycete	-170.8576	0	-205.8572	-135.8581	TRUE
GH-Fungi	PL-Fungi	-256.9688	0	-280.1591	-233.7784	TRUE
GH-Fungi	PL-Oomycete	-240.0799	0	-275.0794	-205.0803	TRUE
GH-Oomycete	GT-Fungi	-79.6354	0	-114.635	-44.6358	TRUE
GH-Oomycete	GT-Oomycete	-77.5556	0	-121.2837	-33.8274	TRUE
GH-Oomycete	PL-Fungi	-163.6667	0	-198.6662	-128.6671	TRUE
GH-Oomycete	PL-Oomycete	-146.7778	0	-190.5059	-103.0497	TRUE
GT-Fungi	GT-Oomycete	2.0799	1	-32.9197	37.0794	FALSE
GT-Fungi	PL-Fungi	-84.0312	0	-107.2216	-60.8409	TRUE
GT-Fungi	PL-Oomycete	-67.1424	0	-102.1419	-32.1428	TRUE
GT-Oomycete	PL-Fungi	-86.1111	0	-121.1107	-51.1115	TRUE
GT-Oomycete	PL-Oomycete	-69.2222	0	-112.9503	-25.4941	TRUE
PL-Fungi	PL-Oomycete	16.8889	0.9101	-18.1107	51.8885	FALSE

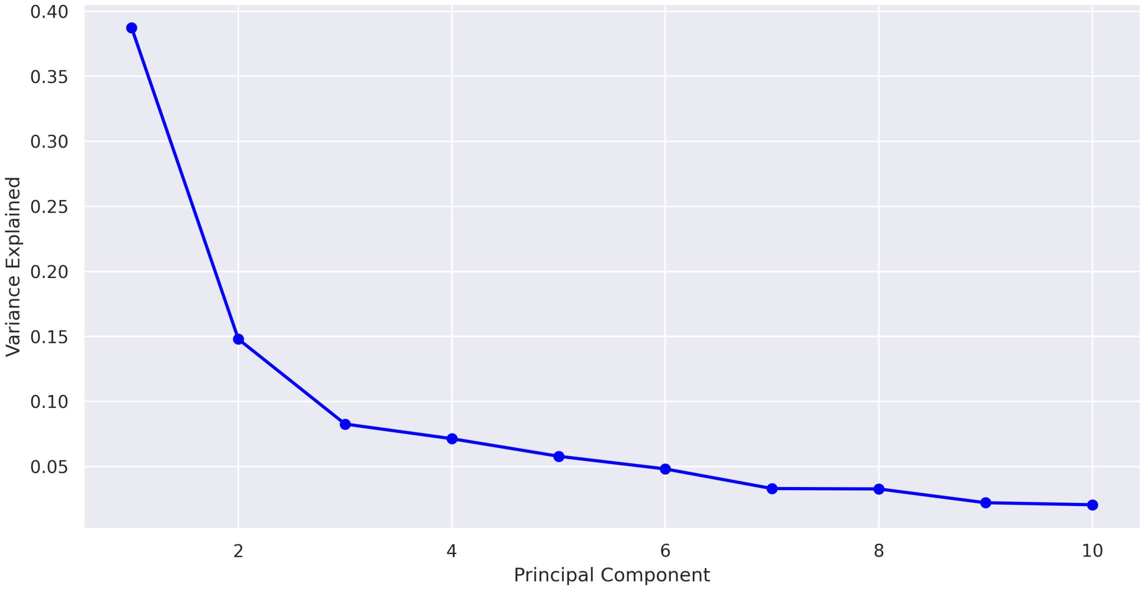
Output from a post hoc TukeyHSD test following a two-way ANOVE testing for statistically significant differences between CAZy class frequencies between fungi and oomycetes.

SI Figure 1: Cumulative explained frequency and scree plot of principal component analysis of CAZyme family frequencies in fungi and oomycetes

CAZyme family frequencies are the number of unique protein IDs assigned to each CAZyme family.



Cumulative explained variance

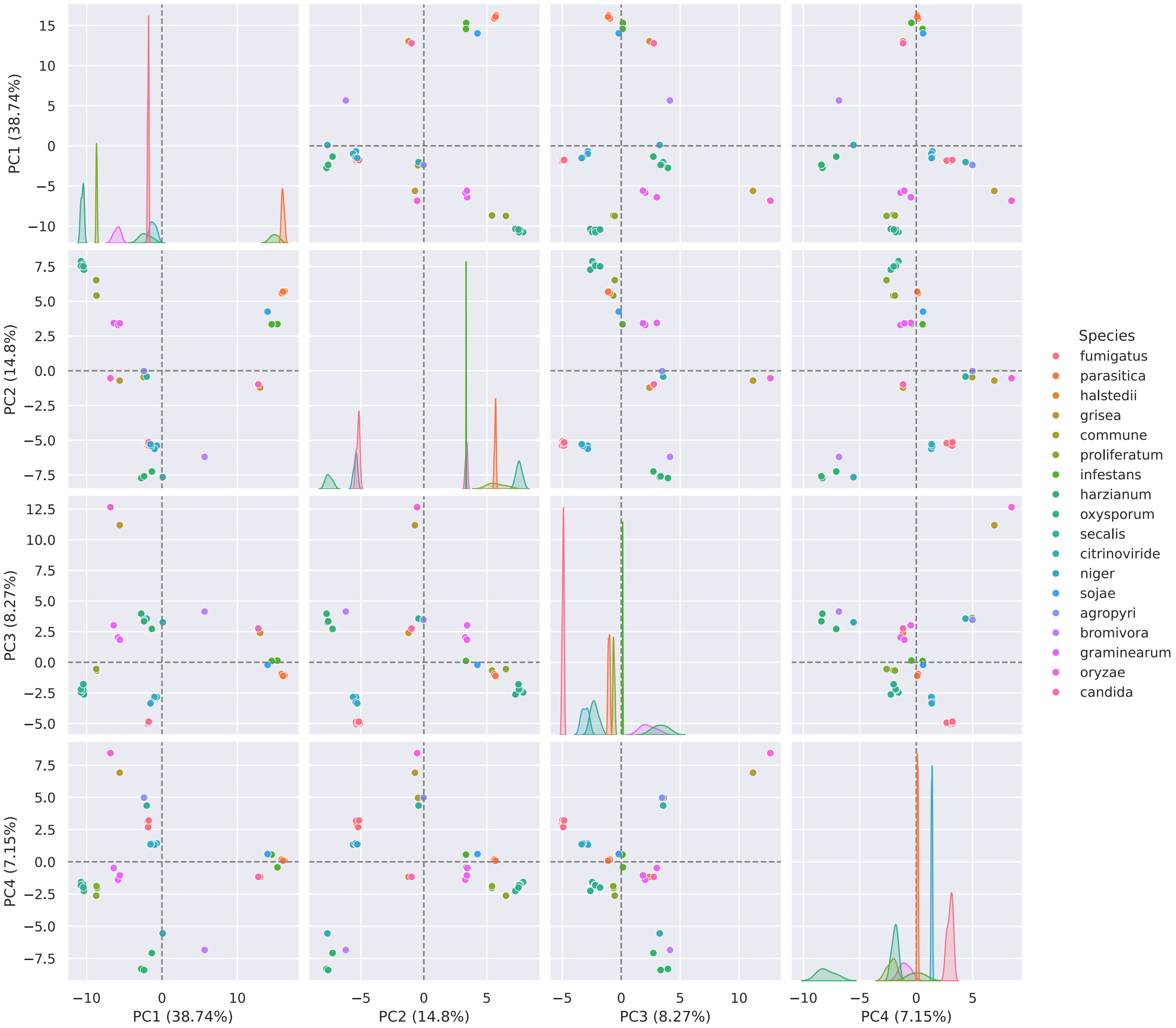


Scree plot

Principal component analysis (PCA) of the CAZy family frequencies in *Pectobacteriaceae* genomes, plotting [A] the cumulative frequency across all computed principal components (PCs). [B] Scree plot plotting the fraction of variance captured by each PC.

SI Figure 2: Principal component analysis of fungal and oomycete CAZyme family frequencies, with genomes colour coded by species

CAZyme family frequencies are the number of unique protein IDs assigned to each CAZyme family.



Principal component analysis of CAZy family frequencies in fungi and oomycete

Genomes are projected onto all combinations of principal components (PCs) PC1-PC4. KDE plots (univariate distribution plots) are plotted on the diagonal, showing the marginal distribution of genomes in each column, colour-coded by species classifications. Scatter plots are colour-coded and styled by species classification.

3 Positive selection screening

SI Table 3: Output from aBSREL for cluster AAA32701

Output from aBSREL for cluster AAA32701. Rows highlighted in green indicate the detection of positive selection

Name	B	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
AAM23009_1	0.172	74.271	0	0	$\omega1 = 0.00$ (81%) $\omega2 = 117$ (19%)
AIE38009_1	0.017	27.042	0	0	$\omega1 = 0.0257$ (99%) $\omega2 = \infty$ (1.3%)
Node29	0.103	17.924	0.001	0	$\omega1 = 0.320$ (93%) $\omega2 = 36.3$ (7.0%)
CAP80630_1	0.043	15.839	0.004	0.0001	$\omega1 = 0.0483$ (96%) $\omega2 = 5000$ (4.1%)
BAB82468_1	0.003	6.4616	0.408	0.0141	$\omega1 = 0.0805$ (100%) $\omega2 = 1500$ (0.37%)
Node6	0.062	5.623	0.604	0.0216	$\omega1 = 0.144$ (95%) $\omega2 = 519$ (4.8%)
Node16	0.068	5.4419	0.639	0.0237	$\omega1 = 0.0754$ (97%) $\omega2 = 5000$ (3.1%)
AAA32701_1	0	0	1	1	$\omega1 = 1.00$ (100%)
AFS18475_1	0	0	1	1	$\omega1 = 1.00$ (100%)
AGV28619_1	0	0	1	1	$\omega1 = 1.00$ (100%)
BAE65949_1	0	0	1	1	$\omega1 = 1.00$ (100%)
BCR99717_1	0	0	1	1	$\omega1 = 1.00$ (100%)
CAK47350_1	0	0	1	1	$\omega1 = 1.00$ (100%)
GAA92866_1	0	0	1	1	$\omega1 = 1.00$ (100%)
GAQ43951_1	0.008	0	1	1	$\omega1 = 0.0295$ (100%)
Node1	0.013	0	1	1	$\omega1 = 0.0292$ (100%)
Node12	1E-04	0	1	1	$\omega1 = 1.00$ (100%)
Node17	0.01	0	1	1	$\omega1 = 0.0664$ (100%)
Node19	0.003	0	1	1	$\omega1 = 0.156$ (100%)
Node22	0.013	3.5707	1	0.0621	$\omega1 = 0.000773$ (97%) $\omega2 = 13.1$ (3.2%)
Node23	0.002	0	1	1	$\omega1 = 0.950$ (100%)
Node24	0	0	1	1	$\omega1 = 0.00$ (100%)
Node31	0	0	1	1	$\omega1 = 1.00$ (100%)
Node4	0.05	0.4531	1	0.3394	$\omega1 = 0.00$ (92%) $\omega2 = 2.10$ (7.6%)
Node5	0.094	3.777	1	0.0558	$\omega1 = 0.0448$ (92%) $\omega2 = \infty$ (7.6%)
Node7	0.002	0	1	1	$\omega1 = 0.00$ (100%)
Node8	0	0	1	1	$\omega1 = 1.00$ (100%)
OXN15357_1	0.171	0	1	1	$\omega1 = 0.00129$ (85%) $\omega2 = 0.517$ (15%)
QMW36737_1	0	0	1	1	$\omega1 = 1.00$ (100%)
QMW48793_1	0	0	1	1	$\omega1 = 1.00$ (100%)
QQK48079_1	0	0	1	1	$\omega1 = 1.00$ (100%)
QRD92196_1	0	0	1	1	$\omega1 = 1.00$ (100%)
SPB48655_1	0	0	1	1	$\omega1 = 1.00$ (100%)

SI Table 4: Output from aBSREL for RKK95495 cluster

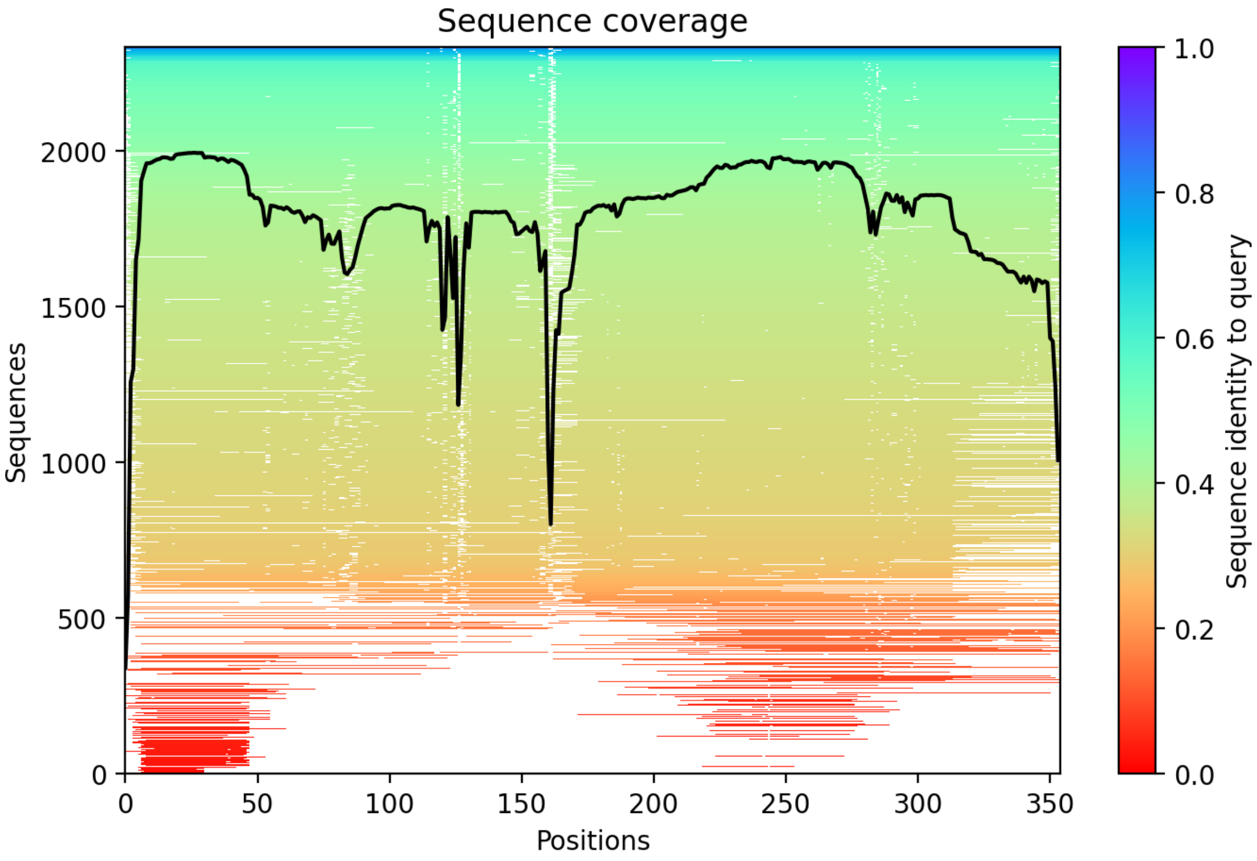
Output from aBSREL for cluster RKK95495. Rows highlighted in green indicate the detection of positive selection

Name	B	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
RKL00490_1	0.0278	30.2322	0	0	$\omega 1 = 0.274$ (98%) $\omega 2 = \infty$ (2.3%)
AAC49420_1	0.1071	10.6468	0.0943	0.0017	$\omega 1 = 0.0594$ (95%) $\omega 2 = 26.0$ (4.6%)
Node41	0.0419	7.6034	0.4326	0.0079	$\omega 1 = 0.656$ (99%) $\omega 2 = \infty$ (0.72%)
CEI68085_1	0.0427	6.4403	0.7677	0.0142	$\omega 1 = 0.0399$ (99%) $\omega 2 = \infty$ (1.2%)
CCT64935_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
CEF86969_1	0.0014	-27.7065	1	0.5	$\omega 1 = 10000000000$ (100%)
CEF87884_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
CEI70384_1	0.0179	0	1	1	$\omega 1 = 0.0134$ (100%)
CZS83373_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
CZS85704_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
Node1	0	0	1	1	$\omega 1 = 1.00$ (100%)
Node10	0.0665	0	1	1	$\omega 1 = 0.133$ (100%)
Node12	0.007	0	1	1	$\omega 1 = 0.00$ (100%)
Node13	0.016	0	1	1	$\omega 1 = 0.00$ (100%)
Node17	0.0813	1.0815	1	0.2344	$\omega 1 = 0.00$ (97%) $\omega 2 = 3.31$ (2.9%)
Node18	0.0464	0	1	1	$\omega 1 = 0.0167$ (100%)
Node20	0.0287	0	1	1	$\omega 1 = 0.0271$ (100%)
Node21	0.021	0	1	1	$\omega 1 = 0.00$ (100%)
Node22	0	-27.662	1	0.5	$\omega 1 = 10000000000$ (100%)
Node23	0	0	1	1	$\omega 1 = 1.00$ (100%)
Node29	0.0955	0	1	1	$\omega 1 = 0.00362$ (100%)
Node30	0.001	0	1	1	$\omega 1 = 0.00$ (100%)
Node31	0	0	1	1	$\omega 1 = 1.00$ (100%)
Node32	0	0	1	1	$\omega 1 = 1.00$ (100%)
Node37	0.0096	0.0264	1	0.4703	$\omega 1 = 10000000000$ (100%)
Node4	0.0155	0	1	1	$\omega 1 = 0.00$ (100%)
Node43	0.0395	4.1898	1	0.0451	$\omega 1 = 0.00$ (99%) $\omega 2 = 17.3$ (0.93%)
Node45	0.0057	0	1	1	$\omega 1 = 0.0819$ (100%)
Node46	0.0189	0	1	1	$\omega 1 = 0.0535$ (100%)
Node47	0.0014	0.6034	1	0.3093	$\omega 1 = 10000000000$ (100%)
Node5	0.0569	0	1	1	$\omega 1 = 0.0304$ (100%)
Node52	0.0214	0	1	1	$\omega 1 = 0.0128$ (100%)
Node55	0.0016	0	1	1	$\omega 1 = 0.00$ (100%)
Node6	0.0544	2.5251	1	0.1073	$\omega 1 = 0.00$ (97%) $\omega 2 = 65.9$ (2.9%)
Node7	0.1663	0	1	1	$\omega 1 = 0.00156$ (94%) $\omega 2 = 0.108$ (5.6%)
Node8	0.1017	4.2612	1	0.0434	$\omega 1 = 0.00$ (93%) $\omega 2 = 23.9$ (6.6%)
PCD34109_1	0.0377	0.4506	1	0.3399	$\omega 1 = 0.00$ (95%) $\omega 2 = 2.14$ (5.1%)
PCD36610_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
QGI61249_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
QGI78430_1	0	-28.6058	1	0.5	$\omega 1 = 10000000000$ (100%)
QGI92147_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
QKD53642_1	0	0	1	1	$\omega 1 = 0.00$ (100%)
QKD61417_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
QPC62756_1	0.0145	0	1	1	$\omega 1 = 0.0723$ (100%)
QPC78206_1	0.0331	0	1	1	$\omega 1 = 0.0135$ (100%)
QPC80127_1	0.0153	0	1	1	$\omega 1 = 0.0646$ (100%)
RBA14607_1	0	0.5816	1	0.3134	$\omega 1 = 10000000000$ (100%)
RBA19176_1	0.0196	0	1	1	$\omega 1 = 0.171$ (100%)
RKK70046_1	0.0026	0	1	1	$\omega 1 = 0.171$ (100%)
RKK75521_1	0.0088	0	1	1	$\omega 1 = 0.0696$ (100%)
RKK88712_1	0	0	1	1	$\omega 1 = 0.00$ (100%)
RKK95495_1	0.0015	0	1	1	$\omega 1 = 0.169$ (100%)
RKK99321_1	0	-27.6767	1	0.5	$\omega 1 = 10000000000$ (100%)
RKL26876_1	0	0	1	1	$\omega 1 = 1.00$ (100%)
RKL28993_1	0.0278	0	1	1	$\omega 1 = 0.00$ (100%)
RKL50686_1	0.0027	0.5786	1	0.314	$\omega 1 = 10000000000$ (100%)
SCO85995_1	0.0042	0	1	1	$\omega 1 = 0.365$ (100%)

4 *In silco* characterisation of a PL1 CAZyme AIE

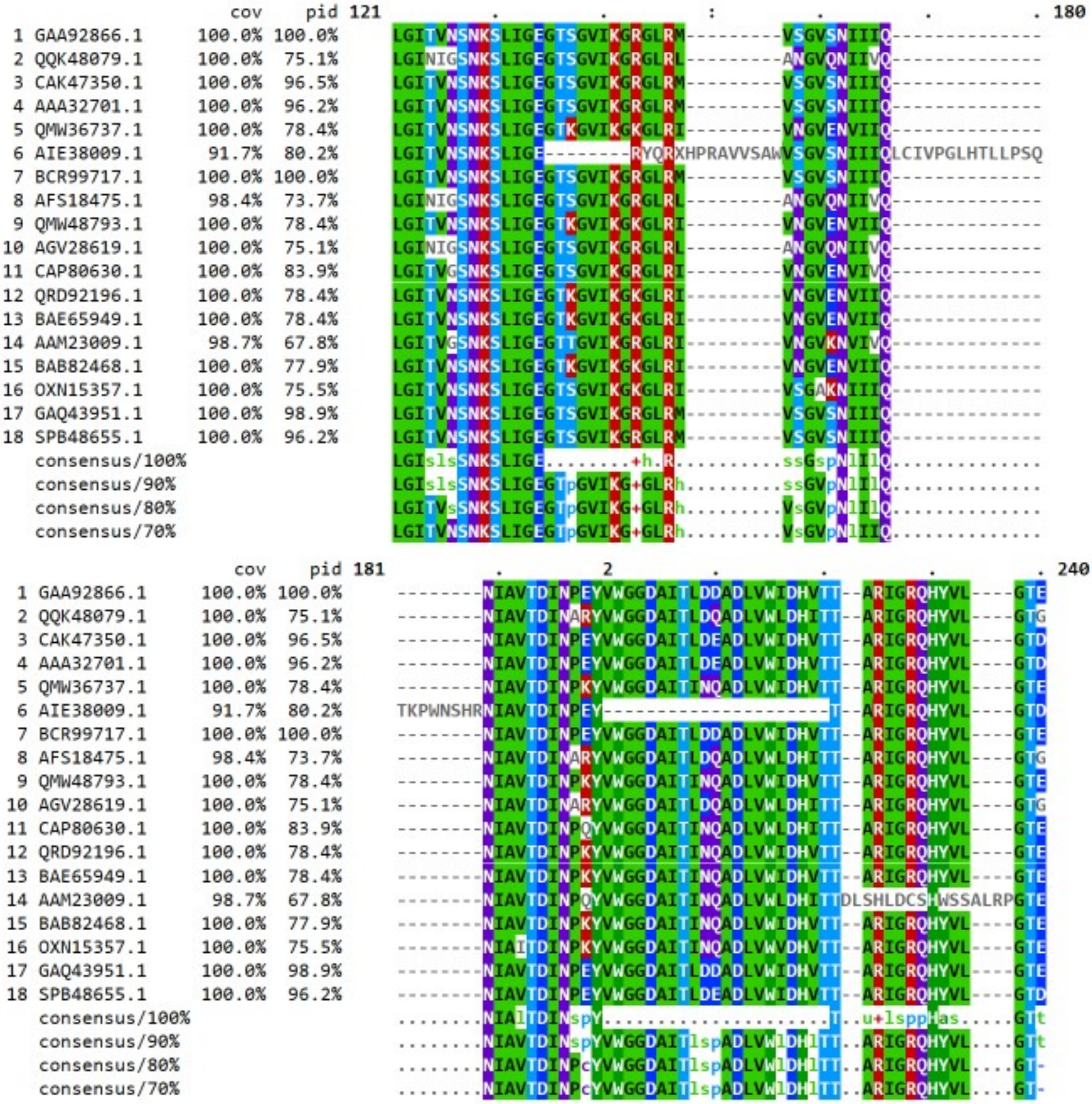
4.1 Prediction of a structural fold for PL1 CAZyme AIE

SI Figure 3: Sequence coverage by ColabFold



Sequence coverage of protein sequence of NCBI:AIE38009.1 by an MSA generated by ColabFold

SI Figure 4: Multi-sequence alignment of AAA32701 cluster members



```
                          cov     pid 121           .         .         :         .         .         . 180
 1 GAA92866.1          100.0% 100.0%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
 2 QQK48079.1          100.0%  75.1%    LGINIGSNKSLIGEGTSGVIKGRGLRL---------ANGVQNIIVQ--------------
 3 CAK47350.1          100.0%  96.5%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
 4 AAA32701.1          100.0%  96.2%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
 5 QMW36737.1          100.0%  78.4%    LGITVNSNKSLIGEGTKGVIKGKGLRI---------VNGVENVIIQ--------------
 6 AIE38009.1           91.7%  80.2%    LGITVNSNKSLIGE--------RYQRXHPRAVVSAWVSGVSNIIIQLCIVPGLHTLLPSQ
 7 BCR99717.1          100.0% 100.0%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
 8 AFS18475.1           98.4%  73.7%    LGINIGSNKSLIGEGTSGVIKGRGLRL---------ANGVQNIIVQ--------------
 9 QMW48793.1          100.0%  78.4%    LGITVNSNKSLIGEGTKGVIKGKGLRI---------VNGVENVIIQ--------------
10 AGV28619.1          100.0%  75.1%    LGINIGSNKSLIGEGTSGVIKGRGLRL---------ANGVQNIIVQ--------------
11 CAP80630.1          100.0%  83.9%    LGITVGSNKSLIGEGTSGVIKGRGLRI---------VNGVENVIVQ--------------
12 QRD92196.1          100.0%  78.4%    LGITVNSNKSLIGEGTKGVIKGKGLRI---------VNGVENVIIQ--------------
13 BAE65949.1          100.0%  78.4%    LGITVNSNKSLIGEGTKGVIKGKGLRI---------VNGVENVIIQ--------------
14 AAM23009.1           98.7%  67.8%    LGITVGSNKSLIGEGTTGVIKGRGLRI---------VNGVKNVIVQ--------------
15 BAB82468.1          100.0%  77.9%    LGITVNSNKSLIGEGTKGVIKGKGLRI---------VNGVENVIIQ--------------
16 OXN15357.1          100.0%  75.5%    LGITVNSNKSLIGEGTSGVIKGKGLRI---------VSGAKNIIIQ--------------
17 GAQ43951.1          100.0%  98.9%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
18 SPB48655.1          100.0%  96.2%    LGITVNSNKSLIGEGTSGVIKGRGLRM---------VSGVSNIIIQ--------------
   consensus/100%                       LGIslsSNKSLIGE........+h.R..........ssGspNlIlQ..............
   consensus/90%                        LGIslsSNKSLIGEGTpGVIKG+GLRh.........ssGVpNlIlQ..............
   consensus/80%                        LGITVsSNKSLIGEGTpGVIKG+GLRh.........VsGVpNlIlQ..............
   consensus/70%                        LGITVNSNKSLIGEGTpGVIKG+GLRh.........VsGVpNlIIQ..............

                          cov     pid 181           .         2         .         .         .         . 240
 1 GAA92866.1          100.0% 100.0%    --------NIAVTDINPEYVWGGDAITLDDADLVWIDHVTT--ARIGRQHYVL----GTE
 2 QQK48079.1          100.0%  75.1%    --------NIAVTDINARYVWGGDAITLDQADLVWLDHITT--ARIGRQHYVL----GTG
 3 CAK47350.1          100.0%  96.5%    --------NIAVTDINPEYVWGGDAITLDEADLVWIDHVTT--ARIGRQHYVL----GTD
 4 AAA32701.1          100.0%  96.2%    --------NIAVTDINPEYVWGGDAITLDEADLVWIDHVTT--ARIGRQHYVL----GTD
 5 QMW36737.1          100.0%  78.4%    --------NIAVTDINPKYVWGGDAITINQADLVWIDHVTT--ARIGRQHYVL----GTE
 6 AIE38009.1           91.7%  80.2%    TKPWNSHRNIAVTDINPEY---------------------T--ARIGRQHYVL----GTD
 7 BCR99717.1          100.0% 100.0%    --------NIAVTDINPEYVWGGDAITLDDADLVWIDHVTT--ARIGRQHYVL----GTE
 8 AFS18475.1           98.4%  73.7%    --------NIAVTDINARYVWGGDAITLDQADLVWLDHITT--ARIGRQHYVL----GTG
 9 QMW48793.1          100.0%  78.4%    --------NIAVTDINPKYVWGGDAITINQADLVWIDHVTT--ARIGRQHYVL----GTE
10 AGV28619.1          100.0%  75.1%    --------NIAVTDINARYVWGGDAITLDQADLVWLDHITT--ARIGRQHYVL----GTG
11 CAP80630.1          100.0%  83.9%    --------NIAVTDINPQYVWGGDAITINQADLVWLDHITT--ARIGRQHYVL----GTE
12 QRD92196.1          100.0%  78.4%    --------NIAVTDINPKYVWGGDAITINQADLVWIDHVTT--ARIGRQHYVL----GTE
13 BAE65949.1          100.0%  78.4%    --------NIAVTDINPKYVWGGDAITINQADLVWIDHVTT--ARIGRQHYVL----GTE
14 AAM23009.1           98.7%  67.8%    --------NIAVTDINPQYVWGGDAITINQADLVWLDHITTDLSHLDCSHWSSALRPGTE
15 BAB82468.1          100.0%  77.9%    --------NIAVTDINPKYVWGGDAITINQADLVWIDHVTT--ARIGRQHYVL----GTE
16 OXN15357.1          100.0%  75.5%    --------NIAITDINPKYVWGGDAITINQADLVWVDHVTT--ARIGRQHYVL----GTE
17 GAQ43951.1          100.0%  98.9%    --------NIAVTDINPEYVWGGDAITLDDADLVWIDHVTT--ARIGRQHYVL----GTE
18 SPB48655.1          100.0%  96.2%    --------NIAVTDINPEYVWGGDAITLDEADLVWIDHVTT--ARIGRQHYVL----GTD
   consensus/100%                       ........NIAlTDINspY.....................T..u+lsppHas.....GTt
   consensus/90%                        ........NIAVTDINspYVWGGDAITlspADLVWlDHlTT..ARIGRQHYVL....GTt
   consensus/80%                        ........NIAVTDINPcYVWGGDAITlspADLVWlDHlTT..ARIGRQHYVL....GT-
   consensus/70%                        ........NIAVTDINPcYVWGGDAITlspADLVWlDHlTT..ARIGRQHYVL....GT-
```

Multi-sequence alignment of AAA32701 cluster members

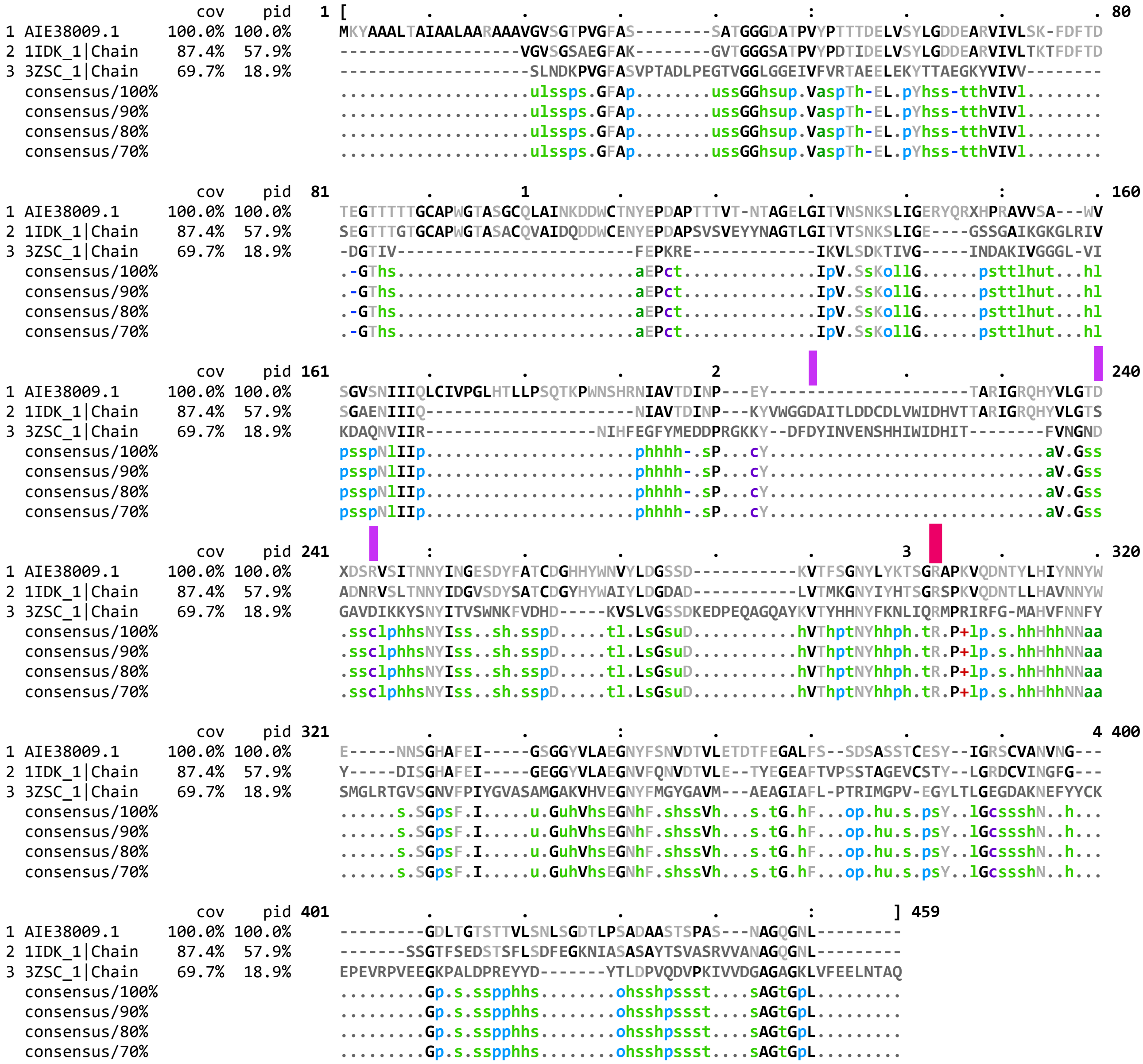
4.2 Structural comparison between AIE, PDB:1IDK and PDB:3ZSC

SI Table 5: Output from Dali server, screening for structural homologs to AIE against the RCSB PDB database (October 2022)

Table 1: Output from Dali server, screening for structural homologs to AIE against the RCSB PDB database (October 2022)

PDB (Chain)	DALI Z-Score	RMSD (Å)	lali	nres	% Identity	Function	Organism	Citation
1idk-(A)	44.3	1.7	170	338	59	Pectin Lyase	Aspergillus niger	Mayans et al., 1997
3zsc-(A)	28.6	2.3	259	329	22	Pectate Lyase	Thermotoga maritima	McDonough et al., 2020
2qy1-(A)	27.6	2.6	257	330	20	Pectate Lyase	Xanthomonas campestris	Xiao et al., 2008

SI Figure 5: Sequence alignment between AIE, PDB:1IDK and PDB:3ZSC



```
                    cov    pid   1 [        .         .         .         .         :         .         .         . 80
1 AIE38009.1     100.0% 100.0%     MKYAAALTAIAALAARAAAVGVSGTPVGFAS--------SATGGGDATPVYPTTTDELVSYLGDDEARVIVLSK-FDFTD
2 1IDK_1|Chain    87.4%  57.9%     -------------------VGVSGSAEGFAK--------GVTGGGSATPVYPDTIDELVSYLGDDEARVIVLTKTFDFTD
3 3ZSC_1|Chain    69.7%  18.9%     --------------------SLNDKPVGFASVPTADLPEGTVGGLGGEIVFVRTAEELEKYTTAEGKYVIVV--------
  consensus/100%                   ....................ulssps.GFAp........ussGGhsup.VaspTh-EL.pYhss-tthVIVl........
  consensus/90%                    ....................ulssps.GFAp........ussGGhsup.VaspTh-EL.pYhss-tthVIVl........
  consensus/80%                    ....................ulssps.GFAp........ussGGhsup.VaspTh-EL.pYhss-tthVIVl........
  consensus/70%                    ....................ulssps.GFAp........ussGGhsup.VaspTh-EL.pYhss-tthVIVl........

                    cov    pid  81          .         1         .         .         .         .         :         . 160
1 AIE38009.1     100.0% 100.0%     TEGTTTTTGCAPWGTASGCQLAINKDDWCTNYEPDAPTTTVT-NTAGELGITVNSNKSLIGERYQRXHPRAVVSA---WV
2 1IDK_1|Chain    87.4%  57.9%     SEGTTTGTGCAPWGTASACQVAIDQDDWCENYEPDAPSVSVEYYNAGTLGITVTSNKSLIGE----GSSGAIKGKGLRIV
3 3ZSC_1|Chain    69.7%  18.9%     -DGTIV-------------------------FEPKRE-------------IKVLSDKTIVG-----INDAKIVGGGL-VI
  consensus/100%                   .-GThs.........................aEPct..............IpV.SsKollG......psttlhut...hl
  consensus/90%                    .-GThs.........................aEPct..............IpV.SsKollG......psttlhut...hl
  consensus/80%                    .-GThs.........................aEPct..............IpV.SsKollG......psttlhut...hl
  consensus/70%                    .-GThs.........................aEPct..............IpV.SsKollG......psttlhut...hl

                    cov    pid 161          .         .         .         2                   .         .           240
1 AIE38009.1     100.0% 100.0%     SGVSNIIIQLCIVPGLHTLLPSQTKPWNSHRNIAVTDINP---EY---------------------TARIGRQHYVLGTD
2 1IDK_1|Chain    87.4%  57.9%     SGAENIIIQ----------------------NIAVTDINP---KYVWGGDAITLDDCDLVWIDHVTTARIGRQHYVLGTS
3 3ZSC_1|Chain    69.7%  18.9%     KDAQNVIIR------------------NIHFEGFYMEDDPRGKKY--DFDYINVENSHHIWIDHIT--------FVNGND
  consensus/100%                   psspNlIIp......................phhhh-.sP...cY.............................aV.Gss
  consensus/90%                    psspNlIIp......................phhhh-.sP...cY.............................aV.Gss
  consensus/80%                    psspNlIIp......................phhhh-.sP...cY.............................aV.Gss
  consensus/70%                    psspNlIIp......................phhhh-.sP...cY.............................aV.Gss

                    cov    pid 241          :         .         .         .         .         3         .         . 320
1 AIE38009.1     100.0% 100.0%     XDSRVSITNNYINGESDYFATCDGHHYWNVYLDGSSD-----------KVTFSGNYLYKTSGRAPKVQDNTYLHIYNNYW
2 1IDK_1|Chain    87.4%  57.9%     ADNRVSLTNNYIDGVSDYSATCDGYHYWAIYLDGDAD-----------LVTMKGNYIYHTSGRSPKVQDNTLLHAVNNYW
3 3ZSC_1|Chain    69.7%  18.9%     GAVDIKKYSNYITVSWNKFVDHD-----KVSLVGSSDKEDPEQAGQAYKVTYHHNYFKNLIQRMPRIRFG-MAHVFNNFY
  consensus/100%                   .ssclphhsNYIss..sh.sspD.....tl.LsGsuD...........hVThptNYhhph.tR.P+lp.s.hhHhhNNaa
  consensus/90%                    .ssclphhsNYIss..sh.sspD.....tl.LsGsuD...........hVThptNYhhph.tR.P+lp.s.hhHhhNNaa
  consensus/80%                    .ssclphhsNYIss..sh.sspD.....tl.LsGsuD...........hVThptNYhhph.tR.P+lp.s.hhHhhNNaa
  consensus/70%                    .ssclphhsNYIss..sh.sspD.....tl.LsGsuD...........hVThptNYhhph.tR.P+lp.s.hhHhhNNaa

                    cov    pid 321          .         .         :         .         .         .         .         4 400
1 AIE38009.1     100.0% 100.0%     E-----NNSGHAFEI-----GSGGYVLAEGNYFSNVDTVLETDTFEGALFS--SDSASSTCESY--IGRSCVANVNG---
2 1IDK_1|Chain    87.4%  57.9%     Y-----DISGHAFEI-----GEGGYVLAEGNVFQNVDTVLE--TYEGEAFTVPSSTAGEVCSTY--LGRDCVINGFG---
3 3ZSC_1|Chain    69.7%  18.9%     SMGLRTGVSGNVFPIYGVASAMGAKVHVEGNYFMGYGAVM---AEAGIAFL-PTRIMGPV-EGYLTLGEGDAKNEFYYCK
  consensus/100%                   ......s.SGpsF.I.....u.GuhVhsEGNhF.shssVh...s.tG.hF...op.hu.s.psY..lGcssshN..h...
  consensus/90%                    ......s.SGpsF.I.....u.GuhVhsEGNhF.shssVh...s.tG.hF...op.hu.s.psY..lGcssshN..h...
  consensus/80%                    ......s.SGpsF.I.....u.GuhVhsEGNhF.shssVh...s.tG.hF...op.hu.s.psY..lGcssshN..h...
  consensus/70%                    ......s.SGpsF.I.....u.GuhVhsEGNhF.shssVh...s.tG.hF...op.hu.s.psY..lGcssshN..h...

                    cov    pid 401          .         .         .         .         :        ] 459
1 AIE38009.1     100.0% 100.0%     ---------GDLTGTSTTVLSNLSGDTLPSADAASTSPAS---NAGQGNL---------
2 1IDK_1|Chain    87.4%  57.9%     -------SSGTFSEDSTSFLSDFEGKNIASASAYTSVASRVVANAGQGNL---------
3 3ZSC_1|Chain    69.7%  18.9%     EPEVRPVEEGKPALDPREYYD-------YTLDPVQDVPKIVVDGAGAGKLVFEELNTAQ
  consensus/100%                   .........Gp.s.sspphhs........ohsshpssst....sAGtGpL.........
  consensus/90%                    .........Gp.s.sspphhs........ohsshpssst....sAGtGpL.........
  consensus/80%                    .........Gp.s.sspphhs........ohsshpssst....sAGtGpL.........
  consensus/70%                    .........Gp.s.sspphhs........ohsshpssst....sAGtGpL.........
```

Figure 2: Protein sequence alignment of AIE, PDB:1IDK and PDB:3ZSC, visualised using MView (at https://www.ebi.ac.uk/Tools/msa/mview/)

5 Positive selection in AIE

SI Table 6: Output from CodeML, detecting sites of positive selection within AIE

Table 2: Output from CodeML Branch site model: Bayes Empirical Bayes (BEB) analysis to detect positive sites for foreground lineages Prob(w¿1) (*P¿0.95, **P¿0.99)

Position	Residue	P-value
67	F	1.000**
68	D	0.999**
69	F	0.999**
70	T	1.000**
71	D	0.999**
72	T	0.999**
73	E	1.000**
74	G	0.999**
75	T	0.999**
80	G	0.998**
81	C	0.965*
82	A	1.000**
83	P	0.886
84	W	1.000**
85	G	0.995**
86	T	1.000**
87	A	0.959*
88	S	0.953*
89	G	0.999**
90	C	0.83
91	Q	1.000**
92	L	0.979*
93	A	0.999**
94	I	0.999**
95	N	0.954*
96	K	0.956*
97	D	0.999**
99	W	1.000**
100	C	0.968*
101	T	1.000**
102	N	0.95
103	Y	0.957*
104	E	0.999**
105	P	1.000**
106	D	1.000**
107	A	0.931
108	P	0.964*
109	T	0.959*
112	V	0.995**
113	T	0.999**
114	N	0.952*
115	T	1.000**
116	A	0.998**
117	G	0.999**
118	E	0.999**
119	L	1.000**
120	G	0.985*
121	I	0.961*
122	T	0.999**
123	V	0.946
124	N	0.952*
125	S	0.999**
127	K	1.000**
128	S	0.999**
129	L	1.000**
130	I	0.999**
131	G	1.000**
132	E	1.000**
133	R	0.997**
134	Y	0.999**
135	Q	1.000**
136	R	1.000**
137	H	1.000**
138	P	1.000**
139	R	0.999**
140	A	1.000**
141	V	0.977*
142	V	0.997**
143	S	1.000**
144	A	1.000**
145	W	0.999**
146	V	0.997**
147	S	0.999**
148	G	0.960*
149	V	0.968*
150	S	0.999**
151	N	0.980*
152	I	0.954*
155	Q	1.000**
156	L	1.000**
157	C	0.999**
159	V	0.973*
160	P	1.000**
161	G	1.000**
162	L	1.000**
163	H	1.000**
164	T	0.975*
165	L	1.000**
166	L	0.999**
167	P	1.000**
168	S	1.000**
169	Q	1.000**
170	T	1.000**
171	K	1.000**
172	P	1.000**
173	W	1.000**
174	N	0.970*
175	S	0.958*
176	H	0.958*
177	R	1.000**
178	N	0.995**
179	I	1.000**
180	A	0.964*
181	V	0.998**
182	T	0.999**
183	D	1.000**
185	N	0.943
186	P	0.956*
187	E	0.998**
188	Y	1.000**
216	F	0.964*

SI Table 7: Output of MEME, detecting positive selection in AIE

Table 3: Sites of positive selection in the protein sequences of cluster AAA32701, detected by MEME

Codon	Partition	alpha	beta+	p+	LRT	Episodic selection detected?	branches	Significance
28	1	0.379	1186.8	0.06	13.1	Yes, p = 0.0006	1	**
32	1	0	7.4	0.21	6	Yes, p = 0.0225	2	*
126	1	0	6.5	0.1	4.5	Yes, p = 0.0494	1	*
144	1	0.366	1889.3	0.06	14.1	Yes, p = 0.0004	1	**
305	1	0	37.3	0.15	7.6	Yes, p = 0.0098	0	**
414	1	0	10.6	0.19	5.2	Yes, p = 0.0340	2	*

6 *In silco* characterisation of a PL3 CAZyme RKL

SI Table 8: Output from CodeML, detecting sites of positive selection within RKL

Table 4: Output from CodeML Branch site model: Bayes Empirical Bayes (BEB) analysis to detect positive sites for foreground lineages Prob(w>1) (*P¿0.95, **P¿0.99)

Position	Residue	P-value	Position	Residue	P-value	Position	Residue	P-value
67	F	1.000**	113	T	0.999**	155	Q	1.000**
68	D	0.999**	114	N	0.952*	156	L	1.000**
69	F	0.999**	115	T	1.000**	157	C	0.999**
70	T	1.000**	116	A	0.998**	159	V	0.973*
71	D	0.999**	117	G	0.999**	160	P	1.000**
72	T	0.999**	118	E	0.999**	161	G	1.000**
73	E	1.000**	119	L	1.000**	162	L	1.000**
74	G	0.999**	120	G	0.985*	163	H	1.000**
75	T	0.999**	121	I	0.961*	164	T	0.975*
80	G	0.998**	122	T	0.999**	165	L	1.000**
81	C	0.965*	123	V	0.946	166	L	0.999**
82	A	1.000**	124	N	0.952*	167	P	1.000**
83	P	0.886	125	S	0.999**	168	S	1.000**
84	W	1.000**	127	K	1.000**	169	Q	1.000**
85	G	0.995**	128	S	0.999**	170	T	1.000**
86	T	1.000**	129	L	1.000**	171	K	1.000**
87	A	0.959*	130	I	0.999**	172	P	1.000**
88	S	0.953*	131	G	1.000**	173	W	1.000**
89	G	0.999**	132	E	1.000**	174	N	0.970*
90	C	0.83	133	R	0.997**	175	S	0.958*
91	Q	1.000**	134	Y	0.999**	176	H	0.958*
92	L	0.979*	135	Q	1.000**	177	R	1.000**
93	A	0.999**	136	R	1.000**	178	N	0.995**
94	I	0.999**	137	H	1.000**	179	I	1.000**
95	N	0.954*	138	P	1.000**	180	A	0.964*
96	K	0.956*	139	R	0.999**	181	V	0.998**
97	D	0.999**	140	A	1.000**	182	T	0.999**
99	W	1.000**	141	V	0.977*	183	D	1.000**
100	C	0.968*	142	V	0.997**	185	N	0.943
101	T	1.000**	143	S	1.000**	186	P	0.956*
102	N	0.95	144	A	1.000**	187	E	0.998**
103	Y	0.957*	145	W	0.999**	188	Y	1.000**
104	E	0.999**	146	V	0.997**	216	F	0.964*
105	P	1.000**	147	S	0.999**			
106	D	1.000**	148	G	0.960*			
107	A	0.931	149	V	0.968*			
108	P	0.964*	150	S	0.999**			
109	T	0.959*	151	N	0.980*			
112	V	0.995**	152	I	0.954*			

SI Table 9: Output of MEME, detecting positive selection in RKL

Table 5: Sites of positive selection in the protein sequences of cluster RKK95495, detected by MEME

Codon	Partition	alpha	beta+	p+	LRT	Episodic selection detected?	branches	Significance
192	1	0	0	0.4	6.98	Yes, p = 0.01	2	**
197	1	0	0	0.09	5.7	Yes, p = 0.03	1	*

SI table 10: Prediction of N-glycosylation in RKL

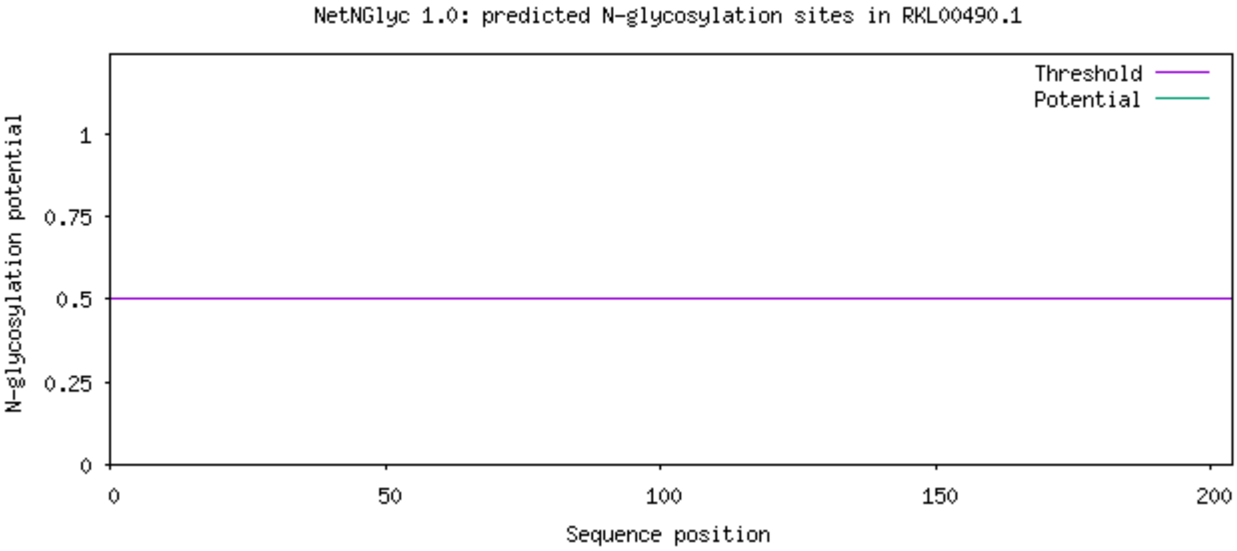


Figure 3: Output from NetNGlyc, predicting sites of N-glycosylation in RKL00490.1

SI table 11: Prediction of O-glycosylation in RKL

Table 6: Output from NetOGlyc, predicting sites of O-glycosylation in RKL00490.1

Sequence name	Start	End	Score	Strand	Frame	Comment
RKL00490_1	5	5	0.370576	.	.	
RKL00490_1	11	11	0.104116	.	.	
RKL00490_1	15	15	0.28624	.	.	
RKL00490_1	23	23	0.645302	.	.	POSITIVE
RKL00490_1	30	30	0.662159	.	.	POSITIVE
RKL00490_1	32	32	0.52429	.	.	POSITIVE
RKL00490_1	34	34	0.683784	.	.	POSITIVE
RKL00490_1	36	36	0.612201	.	.	POSITIVE
RKL00490_1	46	46	0.29102	.	.	
RKL00490_1	61	61	0.203508	.	.	
RKL00490_1	62	62	0.384798	.	.	
RKL00490_1	71	71	0.067379	.	.	
RKL00490_1	99	99	1.76E-05	.	.	
RKL00490_1	101	101	0.003115	.	.	
RKL00490_1	116	116	0.01325	.	.	
RKL00490_1	122	122	0.066702	.	.	
RKL00490_1	162	162	0.022455	.	.	
RKL00490_1	173	173	0.059815	.	.	
RKL00490_1	179	179	0.105596	.	.	
RKL00490_1	180	180	0.213795	.	.	
RKL00490_1	192	192	0.004492	.	.	