

How to bark up the right tree –
Novel metagenomic laccase-like multicopper
oxidases from the European spruce bark beetle
Ips typographus.

Dissertation

to obtain the academic degree of
Doctor rerum naturalium
(Dr. rer. nat.)

at the Department of Biology,
Faculty of Mathematics, Informatics, and Natural Sciences
of the University of Hamburg

Andrea Schorn

from Stade

Hamburg 2020

Approved by the Department of Biology, Faculty of Mathematics, Informatics, and Natural Sciences of the University of Hamburg

at the request of

Prof. Dr. Wolfgang Streit

Second evaluator of the dissertation:

Prof. Dr. Stefan Linder

Date of defense (Disputation):

11. September 2020

This is a corrected version.



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Fakultät für
Mathematik, Informatik
und Naturwissenschaften

UHH
Studienbüro Biologie
Petra Ziemer, Raum CvL/E.522
Ohnhorststr. 18

22609 Hamburg

Prof. Dr. Andrew Torda
Zentrum für Bioinformatik
Bundesstr. 43
20146 Hamburg

Tel. +49 40 42838-7331
Fax +49 40 42838-7332
torda@zbh.uni-hamburg.de
www.zbh.uni-hamburg.de/torda

— 30.03.20

English Language Assessment:

SCHORN, Andrea

"How to bark up...*Ips Typographus*"

By the power vested in me by the elders of the Department of Biology, I pronounce the English in Ms Schorn's thesis to be of scientific standard.

Andrew Torda

Table of Content

Table of Content	I
List of Figures	IV
List of Tables	VII
Abstract	IX
Zusammenfassung	X
1 Introduction	1
1.1 Metagenomics.....	1
1.2 Gut microbiome	1
1.3 The European spruce bark beetle <i>Ips typographus</i>	3
1.4 Multicopper oxidases.....	5
1.4.1 Classification and structure	7
1.4.2 Occurrence and function	11
1.4.3 Industrial fields of application	14
1.5 Aim of the study	15
2 Material und Methods	17
2.1 Bark beetle sample collection and dissection.....	17
2.2 Preparation of buffers with specific pH.....	17
2.3 Bacterial strains, vectors and primers	19
2.4 Cultivation of bacteria.....	22
2.4.1 Antibiotics and supplements.....	22
2.4.2 Determination of cell density	22
2.5 Methods for working with DNA	23
2.5.1 DNA extraction and purification	23
2.5.2 Agarose gel electrophoresis	23
2.5.3 Polymerase chain reaction (PCR)	24
2.5.4 Digestion of DNA by restriction endonucleases.....	26
2.5.5 Ligation of DNA	27
2.5.6 Preparation of chemically competent cells	28
2.5.7 Heat-shock transformation	29
2.5.8 Site-directed mutagenesis.....	29

2.6	Metagenomics	30
2.6.1	Illumina Hiseq sequencing and database analyses	30
2.6.2	Phylogenetic analyses	31
2.7	Protein biochemical methods	32
2.7.1	Expression using auto-induction medium	32
2.7.2	Crude cell extract preparation	33
2.7.3	Protein purification by Fast Protein Liquid Chromatography.....	34
2.7.4	Protein quantification	35
2.7.5	SDS-Polyacrylamide gel electrophoresis (SDS-PAGE).....	35
2.7.6	Enzyme activity assay	37
3	Results	38
3.1	<i>Ips typographus</i> sample collection and dissection	38
3.2	DNA extraction and Illumina HiSeq Sequencing	39
3.3	Phylogenetic analysis by 16S rRNA gene sequencing	40
3.4	Phylogenetic analysis by 18S rRNA gene sequencing	44
3.5	Annotation by the DOE-JGI Microbial Annotation Pipeline.....	45
3.6	Comparison with other gut-associated metagenomic datasets	46
3.7	Novel putative multicopper oxidases	48
3.7.1	Enzymes originating from the <i>Ips typographus</i> dataset.....	48
3.7.2	Enzymes originating from other bacterial strains	52
3.7.3	Phylogenetic comparison of the newfound multicopper oxidases ..	55
3.8	Vectormodification.....	56
3.9	Amplification and cloning of the multicopper oxidase genes	58
3.10	Heterologous gene expression.....	58
3.11	Protein purification.....	60
3.12	Enzyme characterization	64
3.12.1	Structural properties	64
3.12.2	Storage stability	67
3.12.3	Optimum pH	71
3.12.4	Optimum temperature.....	72
3.12.5	Heat stability	72
4	Discussion	76
4.1	Sample collection, DNA extraction and sequencing	76
4.2	Taxonomic distribution and genome size	76

4.3	Phylogenetic analysis by 16S rRNA gene sequencing	77
4.4	Phylogenetic analysis by 18S rRNA gene sequencing	79
4.5	Annotation and comparison of different phylogenetic approaches	80
4.6	Comparison with other gut-associated metagenomic datasets	81
4.7	Novel putative multicopper oxidases	82
4.7.1	Enzymes originating from the <i>Ips typographus</i> dataset	82
4.7.2	Enzymes originating from other bacterial strains	83
4.8	Protein expression and purification	84
4.9	Enzyme characterization	85
4.9.1	Structural properties	85
4.9.2	Storage stability	86
4.9.3	Optimum pH	87
4.9.4	Optimum temperature	87
4.9.5	Heat stability	88
4.10	Conclusion	89
4.11	Outlook	90
	References	91
	List of Abbreviations	102
	Appendix	103
	Declaration on oath / Eidesstattliche Versicherung	118
	Acknowledgements / Danksagung	119

List of Figures

Figure 1:	Imago of the European spruce bark beetle <i>Ips typographus</i>	3
Figure 2:	Reproductive cycle of <i>Ips typographus</i>	4
Figure 3:	The three primary representatives of the MCO enzyme group: laccase, ascorbate oxidase and ceruloplasmin.....	6
Figure 4:	Typical MCO active site and copper ion distribution.	6
Figure 5:	Characteristic blue color of MCOs.	7
Figure 6:	Distribution of known MCO sequences classified into 16 superfamilies	8
Figure 7:	Domain and copper ion organization of several copper-containing enzymes.	9
Figure 8:	Typical structure of laccases and their cupredoxin-like domains with the example of CueO from <i>E. coli</i>	10
Figure 9:	Taxonomic distribution of the MCO sequences in LccED.	11
Figure 10:	Prospective fields of industrial laccase application.	14
Figure 11:	Overview of critical steps this study aimed at.	16
Figure 12:	The type of slot trap and bark beetle pheromone used in this study..	17
Figure 13:	Primer design for site-directed mutagenesis.....	30
Figure 14:	Systematic overview of the site-directed mutagenesis procedure.	30
Figure 15:	<i>Ips typographus</i> fixated on paraffin plates prior to preparation.	38
Figure 16:	Dissection of the bark beetle intestines.	38
Figure 17:	Taxonomic hits distribution of the whole raw <i>Ips typographus</i> metagenomic dataset.	39
Figure 18:	Rarefaction curve of 16S rRNA gene sequences from the <i>Ips typographus</i> gut microbiome.....	40
Figure 19:	Taxonomic classification of the <i>Ips typographus</i> bacterial 16S rRNA gene sequences on phylum and class level.....	42
Figure 20:	Taxonomic classification of the <i>Ips typographus</i> bacterial 16S rRNA gene sequences on genus level.	43
Figure 21:	Rarefaction curve of 18S rRNA gene sequences from the <i>Ips typographus</i> gut microbiome.....	44

Figure 22: Phylogenetic distribution of bacterial genes in the <i>lps typographus</i> dataset.....	45
Figure 23: Comparison of phylogenetic distribution between <i>lps typographus</i> and eight other gut-associated metagenomic datasets.....	47
Figure 24: Comparison of gene functional distribution according to KEGG between <i>lps typographus</i> and eight other gut-associated metagenomic datasets.	48
Figure 25: Phylogenetic tree of the putative <i>lps typographus</i> MCO genes and <i>cueo</i> from <i>E. coli</i>	50
Figure 26: Comparison of the four copper binding motifs of the putative MCOs from the <i>lps typographus</i> metagenome and CueO from <i>E. coli</i>	51
Figure 27: Comparison of the four copper binding motifs of the putative MCOs from <i>Duganella</i> sp. HH01, <i>S. maltophilia</i> K279a and <i>S. fredii</i> NGR234.	54
Figure 28: ORF neighborhood of the putative MCO Genes found by <i>in silico</i> screening of <i>Duganella</i> sp. HH01, <i>S. maltophilia</i> K279a and <i>S. fredii</i> NGR234.....	54
Figure 29: Molecular phylogenetic analysis of the 14 novel MCOs identified in this study in relation to CueO from <i>E. coli</i>	55
Figure 30: Original sequence of pET22b::Strepll in-between the T7 primer binding sites.	57
Figure 31: Site-directed mutagenesis altered sequence of a clone originating from pET22b::Strepll.....	57
Figure 32: Expression of <i>lps</i> 14138 and HH01-482 in T7 SHuffle pre and post auto-induction (A) and comparison of protein yields of <i>lps</i> 1282 and <i>lps</i> 2204 in BL21(DE3) and T7 SHuffle, respectively (B).....	59
Figure 33: Expression of CueO from <i>E. coli</i> , <i>lps</i> 21622 and <i>lps</i> 24328 in BL21(DE3) and T7 SHuffle.....	60
Figure 34: Purification of <i>lps</i> 1282 by Fast Protein Liquid Chromatography.	61
Figure 35: Elution peak of <i>lps</i> 1282 in correlating position to its SDS gel and protein fractions.	62
Figure 36: Gradual migration of the protein front during elution of <i>lps</i> 14138.	63
Figure 37: SDS gel of the purification of <i>lps</i> 14138.....	63
Figure 38: Template and model of the potential 3D structure of HH01-482.....	65
Figure 39: Template and models of the tertiary structures of <i>lps</i> 2204, <i>lps</i> 14138 and <i>lps</i> 24328.....	67

Figure 40: Storage stability of Ips2204 under different temperature and glycerol conditions.....	68
Figure 41: Storage stability of Ips14138 under different temperature and glycerol conditions.....	69
Figure 42: Storage stability of Ips24328 under different temperature and glycerol conditions.....	69
Figure 43: Storage stability of HH01-482 under different temperature and glycerol conditions.....	70
Figure 44: Optimum pH for the activity of Ips2204, Ips14138, Ips24328 and HH01-482 on ABTS.	71
Figure 45: Optimum temperature for the activity of Ips2204, Ips14138, Ips24328 and HH01-482 on ABTS.	72
Figure 46: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 95 °C.....	73
Figure 47: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 80 °C.....	74
Figure 48: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 65 °C.....	74
Figure 49: Vector map of pET22b::StreptII	103
Figure 50: Vector map of pET22b-SDM::StreptII	103
Figure 51: Comparative enzyme activity of three MCOs on ABTS, syringaldazine and guaiacol.	105

List of Tables

Table 1:	Classification of MCOs.....	8
Table 2:	Preparation of citrate-phosphate buffer with specific pH.....	18
Table 3:	Preparation of acetate buffer 0.1 M with specific pH.....	18
Table 4:	Preparation of sodium phosphate buffer 0.1 M with specific pH	18
Table 5:	Bacterial strains used in this thesis	19
Table 6:	Vectors used in this thesis	19
Table 7:	Plasmids used and constructed in this thesis	19
Table 8:	Primer used in this thesis.....	20
Table 9:	Used antibiotics, their stock and applied concentrations	22
Table 10:	TAE buffer (50x)	24
Table 11:	Loading dye	24
Table 12:	Standard recipe for 100 µl of PCR mix	25
Table 13:	Standard PCR program	25
Table 14:	Standard recipe for 50 µl of Phusion PCR mix.....	26
Table 15:	Phusion PCR program.....	26
Table 16:	Analytical and preparative reaction mixture for the digestion of DNA by restriction endonucleases	27
Table 17:	Standard ligation reaction	27
Table 18:	Composition of a pDrive ligation mixture	28
Table 19:	TFB1 buffer.....	29
Table 20:	TFB2 buffer.....	29
Table 21:	ZY medium	33
Table 22:	M solution (50x)	33
Table 23:	5052 solution (50x)	33
Table 24:	Auto-induction medium	33
Table 25:	StrepTrap™ binding buffer (PBS).....	34

Table 26: StrepTrap™ elution buffer.....	34
Table 27: StrepTrap™ regeneration buffer	34
Table 28: Standard recipe for a 12 % SDS-PAGE gel	36
Table 29: Resolving gel buffer (4-fold).....	36
Table 30: Stacking gel buffer (4-fold).....	36
Table 31: Electrophoresis buffer (10-fold).....	37
Table 32: SDS loading buffer (4-fold)	37
Table 33: Coomassie stain	37
Table 34: Destaining solution	37
Table 35: Taxonomic distribution of <i>Ips typographus</i> metagenome 16S rRNA gene OTUs.	41
Table 36: Significant <i>Ips typographus</i> metagenome 18S rRNA gene OTUs.....	44
Table 37: Putative MCO genes found by in silico screening of the <i>Ips typographus</i> metagenome.....	49
Table 38: BLAST results and potential origin of the putative <i>Ips typographus</i> MCO genes.....	49
Table 39: Putative MCO genes found by in silico screening of several bacterial genomes.....	52
Table 40: Comparison of experimental steps achieved for each novel putative MCO gene	64
Table 41: Classification of the novel MCO protein sequences.....	65
Table 42: Phyla comparison of the <i>Ips typographus</i> 16S rRNA sequences and all annotated protein-coding genes.	80
Table 43: Raw data of <i>Ips typographus</i> 16S rRNA sequences.....	104
Table 44: Phylogenetic distribution of protein-coding genes from several metagenomic datasets.....	104
Table 45: Functional distribution of protein-coding genes from several metagenomic datasets.....	105

Abstract

Laccases and laccase-like multicopper oxidases (MCOs) – especially from bacterial origin – represent a massive metabolic potential for a large variety of possible industrial applications. For instance, lignin represents a significant portion of plant biomass that is essentially unutilized at present. The ability to efficiently utilize this biopolymer can possibly enable a renewable replacement for fossil-based resources in the future.

As various insects and in particular wood-feeding beetles were shown to exhibit laccase-like activity in their digestive systems, the European spruce bark beetle *Ips typographus* was chosen for metagenomic sequencing and to scavenge for yet unknown MCOs. Phylogenetic analyses revealed limited microbial diversity and a dominance of Proteobacteria with *Morganella* being the most prominent bacterial genus. In addition, eukaryotic organisms like the nematode *Mycoletzkyia buetschlii* – exclusively known to associate with *Ips typographus* – and a potential symbiotic yeast of the genus *Wickerhamomyces* were also discovered.

A total of 219 Mbp was assembled with a number of 127909 unique predicted ORFs. Following upload and annotation, six novel enterobacterial MCOs were identified in the *Ips typographus* dataset *in silico*. Along with one MCO originating from *Duganella* sp. HH01, three of these metagenomic candidates were successfully cloned, heterologously expressed in *E. coli* and characterized regarding their storage, pH and temperature properties. Despite their striking structural resemblance, these three enzymes displayed huge differences, especially in their reaction to high temperatures. Initially, Ips14138 was thought to be the least promising due to its close phylogenetic relation to CueO, a known MCO playing an important role in copper detoxification in *E. coli*. Surprisingly, it was discovered to exhibit an exceptional heat stability and a strong thermal activation effect unknown for CueO but previously detected in MCOs from *Bacillus* origin. Combined with its excellent storage stability, Ips14138 fulfills essential requirements for industrial utilization making it a promising candidate for further research. In conclusion, the gut of the bark beetle *Ips typographus* not only presents a remarkable source for novel metagenomic MCOs, but especially regarding various metabolic functions, there is still a vast enzymatic potential waiting to be explored.

Zusammenfassung

Laccasen und laccaseartige Multi-Kupfer-Oxidase (MCOs) – insbesondere aus Bakterien – bergen ein enormes Stoffwechselfpotenzial für eine große Vielfalt an möglichen industriellen Anwendungen. So stellt beispielsweise Lignin einen bedeutenden Teil der pflanzlichen Biomasse dar, der derzeit im Wesentlichen ungenutzt ist. Die Fähigkeit, dieses Biopolymer effizient nutzen zu können, kann künftig einen erneuerbaren Ersatz für fossile Ressourcen bieten.

Da bei verschiedenen Insekten und insbesondere bei holzfressenden Käfern Laccase-Aktivität im Verdauungssystem nachgewiesen wurde, fiel die Wahl für Metagenomsequenzierung und die Suche nach noch unbekanntem MCOs auf den europäischen Fichtenborkenkäfer *Ips typographus*. Phylogenetische Analysen ergaben eine limitierte mikrobielle Diversität und eine Dominanz der Proteobakterien, wobei *Morganella* die bedeutendste bakterielle Gattung darstellte. Darüber hinaus wurden eukaryotische Organismen wie beispielsweise der Nematode *Mycoletzkyia buetschlii*, von dem bekannt ist, dass er ausschließlich mit *Ips typographus* assoziiert ist, und eine potenziell symbiotische Hefe der Gattung *Wickerhamomyces* entdeckt.

Insgesamt wurden 219 Mbp zusammengestellt, wobei eine Anzahl von 127909 einzigartigen ORFs vorhergesagt wurden. Nach Upload und Annotation wurden sechs neue enterobakterielle MCOs im *Ips typographus*-Datensatz *in silico* identifiziert. Zusammen mit einer MCO aus *Duganella* sp. HH01 wurden drei dieser metagenomischen Kandidaten erfolgreich kloniert, heterolog in *E. coli* exprimiert und hinsichtlich ihrer Lagerungs-, pH- und Temperatureigenschaften charakterisiert. Trotz ihrer auffallenden strukturellen Ähnlichkeit zeigten diese drei Enzyme enorme Unterschiede, insbesondere bei der Reaktion auf hohe Temperaturen. Anfangs wurde Ips14138 aufgrund seiner engen phylogenetischen Verwandtschaft zu CueO, einer bekannten MCO, die eine wichtige Rolle bei der Kupferentgiftung in *E. coli* spielt, als das am wenigsten vielversprechende Enzym angesehen. Es wurde erstaunlicherweise entdeckt, dass es eine außergewöhnliche Hitzestabilität und einen starken thermischen Aktivierungseffekt aufwies, der für CueO bisher unbekannt war, aber zuvor bereits bei MCOs aus diversen *Bacillus*-Spezies festgestellt. Kombiniert

mit seiner ausgezeichneten Lagerfähigkeit, erfüllt Ips14138 wesentliche Anforderungen für eine industrielle Nutzung und ist damit ein ausgezeichneter Kandidat für weitere Forschungsziele.

Zusammenfassend stellt der Datensatz des Borkenkäfers *Ips typographus* nicht nur eine außergewöhnliche Quelle für neuartige metagenomische MCOs dar, sondern vor allem in Bezug auf verschiedenste Stoffwechselfunktionen wartet darin noch ein enormes enzymatisches Potenzial darauf, erforscht zu werden.

1 Introduction

1.1 Metagenomics

With traditional microbiological methods, only a very small fraction of the bacterial diversity (<1 %) can be isolated and cultivated under laboratory conditions. Therefore, a culture-independent approach to identify and study the bigger part of the prokaryotic population symbolizes a groundbreaking revolution in microbiology (Amann *et al.* 1995; Alneberg *et al.* 2018). The enormous size of prokaryotic populations together with an extremely high growth rate represent a large capability for mutations and therefore genetic diversity (Whitman *et al.* 1998). Novel molecular methods focus directly on the whole DNA of an environmental sample – the “metagenome” – to access a massive, formerly untapped potential (Handelsman *et al.* 1998). Initial metagenomic approaches included plasmid and fosmid library construction and screening. Newer, less biased and time-consuming methods like Next Generation Sequencing (NGS) gradually take their place and enable an increase in sequencing depth and data size (Vollmers *et al.* 2017).

Beside new advances in basic research of microbial ecology, metagenomics enables the discovery of genes that code for novel and unique proteins. Especially promising for this purpose are environments with an abundance of divers and strong metabolic activity (Krishnan *et al.* 2014; Hiraoka *et al.* 2016). Due to their higher abundance of lignocellulolytic enzymes, natural habitats like the gastrointestinal tracts of herbivores may be more favorable compared to anthropogenic ecosystems like biogas reactors (Güllert *et al.* 2016). These host-associated systems also show a lot of promise for the discovery of novel laccases as lignin constituents and plant phenolics appear to be part of these enzymes native substrate range (Cañas & Camarero 2010).

1.2 Gut microbiome

In search of novel and untapped sources for industrially applicable enzymes, scientists and research companies increasingly contemplate the massive biotechnological potential of gut-associated microbiota either in respect of quantity or quality

(Krishnan *et al.* 2014). For example, a large diversity in lignocellulolytic genes has been detected in the fecal metagenome of the Asian elephant (Ilmberger *et al.* 2014), while there was only one singular lignolytic enzyme evident in the gastrointestinal tract of the giant panda which seems to be responsible for the digestion of bamboo lignin (Fang *et al.* 2012).

Insect digestive tracts represent a major genetic reservoir as their structures and environmental conditions differ extremely among species. Confined microorganisms provide a broad variety of beneficial factors to their host who potentially gains nutrition, communication and protection against parasites and pathogens. Besides bacteria, symbiotic yeasts also play a role in the insect microbiome in aiding in foraging, digestion or in being a source of nutrition themselves. In comparison to mammalian guts, the number of different microbial species is restricted but highly specialized (Engel & Moran 2013; Stefanini 2018).

Cross-taxon analyses showed that insect-associated bacterial communities typically encompass only up to 8 phylotypes with the most abundant one accounting for more than 50 % of the whole microbiome on average. Almost 70 % of these phylotypes were limited to unique insect species. In the majority of analyzed samples, Proteobacteria were primarily contributing to the dominant taxa (Jones *et al.* 2013).

It was previously discovered that numerous bacteria underwent a significant genomic decay during adaptation to symbiotic lifestyle. They lost genes futile for their new environment and established common symbiont characteristics like extremely fast protein evolution and a high abundance of chaperones (Burke & Moran 2011; McCutcheon & Moran 2011).

In Particular, xylophagous beetles like the Asian longhorned beetle *Anoplophora glabripennis* were reported to have lignin degrading capability during passage through the gut (Geib *et al.* 2008; Engel & Moran 2013). Putative bacterial and fungal genes for laccases, multicopper oxidases (MCOs) and polyphenol oxidases were annotated in the corresponding metagenome, hinting at a large potential for lignin degradation (Scully *et al.* 2013). Another wood-feeding beetle *Odontotaenius disjunctus* possesses a digestive tract divided in four compartments with different anatomical properties, environmental conditions like oxygen concentration and therefore distinct microbiomes. These specific host environments and nutritional habits promote metabolic functions that are key for lignocellulose deconstruction

(Ceja-Navarro *et al.* 2019). Lignin degrading enzymes are especially interesting because a large portion of plant biomass is virtually unutilized and can potentially be used as a renewable replacement for fossil-based resources in the future (Christopher *et al.* 2014; Chauhan *et al.* 2017).

Overall, the gut metagenome of a wood-feeding beetle appears to be a promising source for novel lignolytic enzymes.

1.3 The European spruce bark beetle *Ips typographus*

The European spruce bark beetle *Ips typographus* (Figure 1) belongs to the Scolytinae subfamily of Coleoptera. It is one of the major forest pests in Central Europe and poses a recurrent threat to its main host tree, the Norway spruce *Picea abies* (Christiansen & Bakke 1988).

Both larva and imago feed on the bast layer, a part of the phloem tissue. The beetles reproduce under the bark and create tunnels – so called galleries – that may lead to the disruption of the trees nutrient transport (Lobinger *et al.* 2016). The adult beetles are dark brown in color, 4.2 to 5.5 cm large and preferably infest weakened or dying spruce trees and fallen logs (NW-FVA 2009). If the quantity of attacking beetles is high enough, they can overcome the trees defensive system and kill previously healthy specimens (Schroeder 2001).



Figure 1: Imago of the European spruce bark beetle *Ips typographus* (Altmann 2008).

Once beetles colonize the host tree, they produce pheromones that are distributed by emerging wood dust and attract other mature beetles over a large distance. When the population reaches a certain threshold density, they also produce repellants that redirect arriving beetles to nearby trees (NW-FVA 2009).

The swarming and therefore the start of the reproductive cycle of *Ips typographus* (Figure 2) is highly dependent on sunshine and continuous warmth. Following hibernation, imagos of this bark beetle initiate the colonization process at air temperatures above 16.5 °C. Male beetles set off earlier than females. So-called pioneer beetles discover and infest suitable breeding material (Wermelinger 2004).

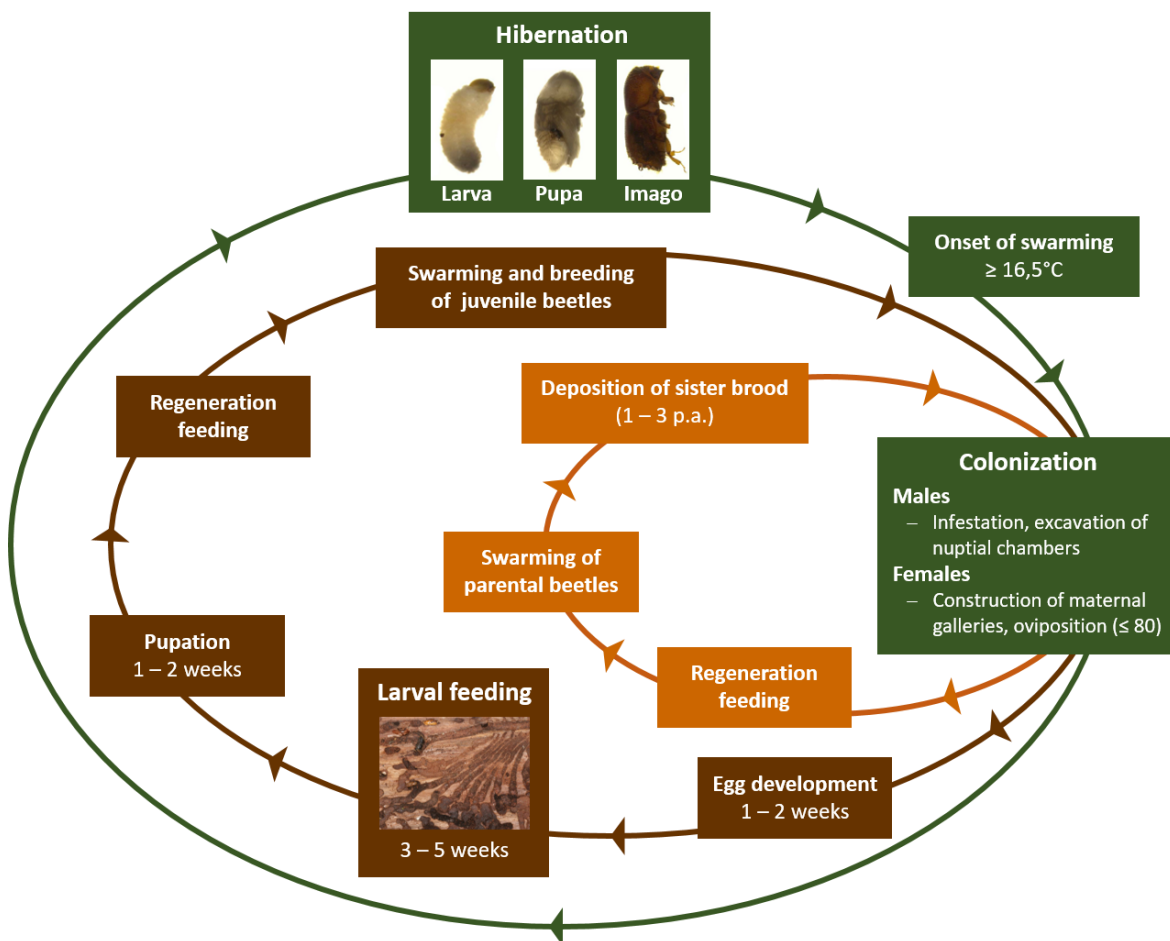


Figure 2: Reproductive cycle of *Ips typographus*. The European spruce bark beetle *Ips typographus* hibernates in larval, pupal, or adult state. With temperatures of 16.5 °C and above they initiate the swarming and colonization process. Male beetles start infestation and excavate nuptial chambers following a successful intrusion under the bark. Afterwards, females construct maternal galleries and deposit up to 80 eggs. Following a regeneration feeding, they swarm again and deposit a sister brood in another host tree. Depending on climate characteristics, they can breed up to 3 times per year. Before larval hatching, the egg development takes 1 to 2 weeks. After 3 to 5 weeks of feeding on the bast layer of the host tree, the larvae pupate for 1 to 2 weeks. The new developed juvenile beetles also undergo a regeneration feeding before they swarm and breed themselves (adapted after (Wermelinger 2004; NW-FVA 2009; Lobinger *et al.* 2016); pictures: (Altmann 2005; John 2016)).

After drilling through the bark, they excavate nuptial chambers which female beetles use to start the construction of maternal galleries. They lay up to 80 eggs and undergo a regeneration feeding before they emerge again to deposit a sister brood at another host tree. Depending on the climate, they can breed up to 3 times per year (Wermelinger 2004; NW-FVA 2009; Lobinger *et al.* 2016).

Egg maturation takes 1 - 2 weeks before the larvae hatch. After 3 - 5 weeks of feeding on their host trees, they pupate for 1 - 2 weeks and transform into light brown juvenile beetles. These also undergo a regeneration feeding and emerge from the bark to swarm and breed themselves. If the weather stays beneficial for the reproduction of *Ips typographus* (Figure 2), up to 3 generations of beetles and sister broods can originate from one female per year (NW-FVA 2009; Lobinger *et al.* 2016).

Literature of microbiological research of Scolytinae mostly consist of taxonomic studies by means of 16S rRNA and not of the construction of whole metagenomic datasets. This comprehensive approach would also enable an in-depth data mining for different classes of target enzymes and numerous basic research subjects in the future. Insects in general possess a highly specialized microbiome that is widely unique for each species (Engel & Moran 2013; Jones *et al.* 2013). Furthermore, symbiotic microorganisms show an extremely accelerated protein evolution, which increases the likelihood of discovering exceptional and yet unidentified enzymes (Burke & Moran 2011; McCutcheon & Moran 2011).

Therefore, the European spruce bark beetle *Ips typographus* appears to be an especially promising source for various kinds of unknown enzymes – especially with lignolytic properties – as different xylophagous beetles exhibit laccase-like activity in their digestive systems (Geib *et al.* 2008).

1.4 **Multicopper oxidases**

Multicopper oxidases (MCO) are blue copper-containing enzymes that oxidize their substrate while reducing dioxygen to water. These enzymes possess significant advantages over other oxidases because they are highly stable and their sole byproduct is water (Reiss *et al.* 2013). The three main representatives of this enzyme group are laccases (EC 1.10.3.2) (Figure 3A), ascorbate oxidases (AO) (EC 1.10.3.3) (Figure 3B) and ceruloplasmin (EC 1.16.3.1) (Figure 3C) (Messerschmidt 1997).

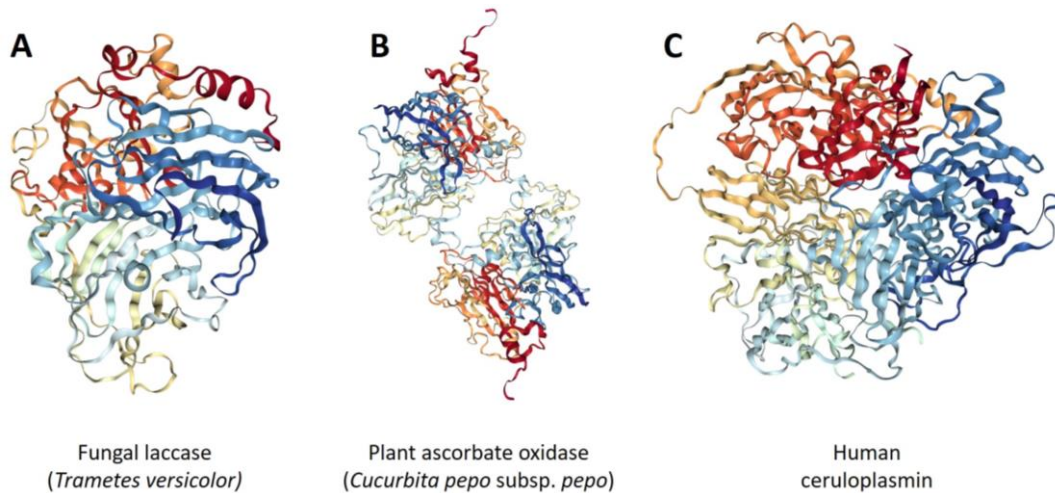


Figure 3: The three primary representatives of the MCO enzyme group: laccase, ascorbate oxidase and ceruloplasmin. The following images of the molecular structures in rainbow color scheme originate from the protein data bank (PDB) (Berman *et al.* 2000) and were visualized with the NGL viewer (Rose *et al.* 2018): *T. versicolor* laccase (A) (PDB ID: 1GYC) (Piontek *et al.* 2002), *C. pepo* subsp. *pepo* ascorbate oxidase (B) (PDB ID: 1AOZ) (Messerschmidt *et al.* 1992) and human ceruloplasmin (C) (PDB ID: 4ENZ) (Samyгина *et al.* 2013).

All MCOs possess at least one type 1 (T1), one type 2 (T2) and two type 3 (T3) copper ions that are allocated by their different spectroscopic properties. These ions are situated at two active sites, the T1 copper ion is located in the mononuclear copper center (MNC) where the substrate is oxidized and the T2/T3 copper ions shape the trinuclear copper center (TNC) where the reduction of dioxygen takes place (Figure 4) (Reiss *et al.* 2013). The positions of these copper ions are maintained by various amino acid (AA) side chains, primarily histidine. One cysteine and two histidine residues hold the T1 copper ion in place while one isoleucine, leucine or methionine is stationed nearby.

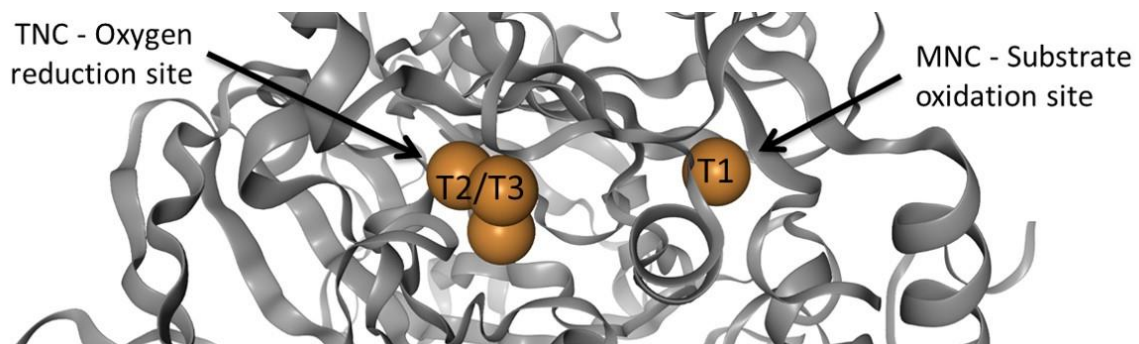


Figure 4: Typical MCO active site and copper ion distribution. A minimum of four copper ions (orange orbs) are separated into two active sites. The type-1 copper ion (T1) is located at the mononuclear copper center (MNC) where the substrate oxidation takes place while one type-2 (T2) and two type-3 (T3) copper ions shape the trinuclear copper center (TNC) where the dioxygen is reduced at. The image shows a section of the protein structure of CueO from *E. coli* (PDB ID: 3PAV) (Montfort *et al.* 2011).

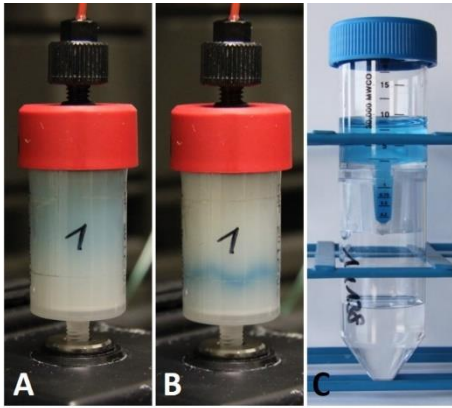


Figure 5: Characteristic blue color of MCOs. Images of the novel blue MCO Ips14138 identified and characterized throughout this study: **A** and **B** during fast protein liquid chromatography (FPLC) and **C** in purified and concentrated solution.

Due to an intense absorption at 614 nm wavelength caused by the highly covalent bond between Cu^{2+} and cysteine, the MNC induces the enzyme groups characteristic blue color (Figure 5) (Enguita *et al.* 2003). The T2 copper is bound by two histidine residues whereas the two T3 copper ions are organized by six histidine residues in total. In summary, these copper ligands compose four AA motifs that are highly conserved in MCOs: H-x-H-G, H-x-H, H-x(2)-H-x-H and H-C-H-x(3)-H-x(4)-[MFL] (Giardina *et al.* 2010; Reiss *et al.* 2013). These motifs can be used to screen genomic and metagenomic datasets for yet unknown MCOs to utilize this still vastly untapped enzymatic potential.

1.4.1 Classification and structure

The enzyme group of MCOs is currently classified into 16 superfamilies and 105 homologous families in the biocatnet Laccase and Multicopper Oxidase Engineering Database LccED (Sirim *et al.* 2011; Buchholz *et al.* 2016).

This database is designated for systematic, sequence-based classification and analysis of the MCO protein family. It currently contains the data of 15017 protein sequences, 14537 of them linked to their respective superfamily (Figure 6), and 138 enzyme structures transmitted from the National Center for Biotechnology Information NCBI database (<https://www.ncbi.nlm.nih.gov>) and the Protein Data Bank PDB (Berman *et al.* 2000) (Table 1). Based on number of identified structures alone, the Basidiomycete Laccase superfamily is the second best explored. Only the bacterial copper efflux oxidase (CueO) type MCOs, which account for roughly 19 % of sequences and $\frac{1}{3}$ of all MCO structures in the LccED were researched more. In short, bacterial MCOs contribute to more than half of all available protein sequences in the LccED (biocatnet LccED v6.4: <https://lcced.biocatnet.de/> access 2019-07-21).

Table 1: Classification of MCOs (according to LccED*)

#	Superfamily	Group	Published Sequences	Published Structures
1	A - Basidiomycete Laccase	3dMCO	924 (6.4 %)	29
2	B - Ascomycete MCO	3dMCO	696 (4.8 %)	12
3	C - Insect Laccase	3dMCO	286 (2.0 %)	0
4	D - Fungal Pigment MCO	3dMCO	282 (1.9 %)	0
5	E - Fungal Ferroxidase	3dMCO	530 (3.6 %)	1
6	F - Fungal and Plant AO	3dMCO	672 (4.6 %)	4
7	G - Plant Laccase	3dMCO	1395 (9.6 %)	0
8	H - Bacterial CopA	3dMCO	1877 (12.9 %)	0
9	I - Bacterial Bilirubin Oxidase	3dMCO	1161 (8.0 %)	19
10	J - Bacterial CueO	3dMCO	2781 (19.1 %)	47
11	K - SLAC-like type B	2dMCO	237 (1.6 %)	14
12	L - Bacterial MCO	3dMCO	2001 (13.8 %)	1
13	M - Archaeal type A	2dMCO	63 (0.4 %)	0
14	N - Bacterial type B	2dMCO	641 (4.4 %)	0
15	O - Archaeal and Bacterial type C	2dMCO	366 (2.5 %)	7
16	P - Ceruloplasmin	6dMCO	625 (4.3 %)	4
In total			14537 (100 %)	138

*biocatnet LccED v6.4: <https://lcced.biocatnet.de/> access 2019-07-21 (Sirim *et al.* 2011; Buchholz *et al.* 2016)

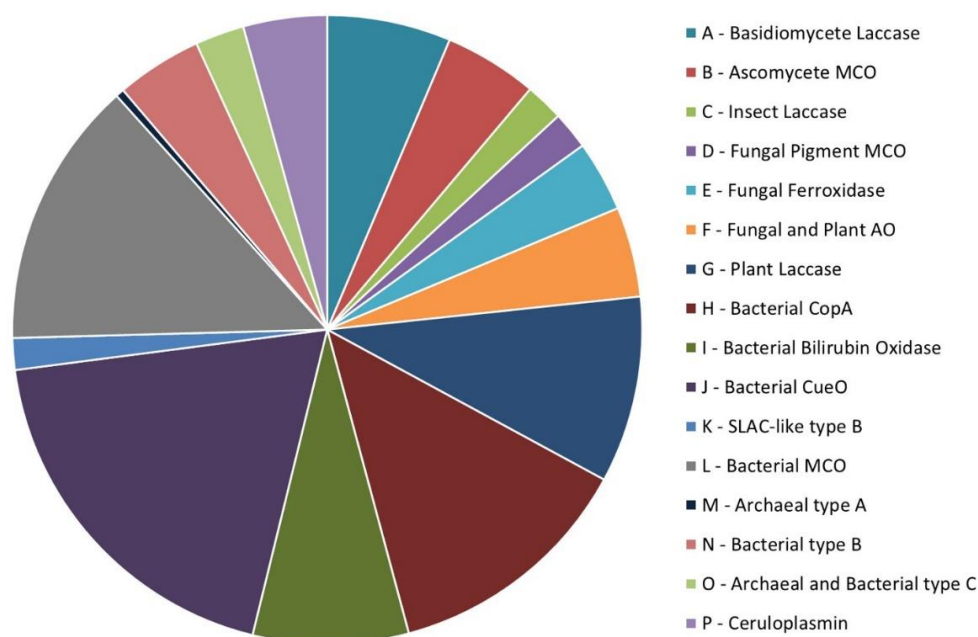


Figure 6: Distribution of known MCO sequences classified into 16 superfamilies according to biocatnet LccED (v6.4: <https://lcced.biocatnet.de/> access 2019-07-21). (Sirim *et al.* 2011; Buchholz *et al.* 2016) The respective percentages are shown in Table 1.

MCOs are also classifiable by their number of domains, as small laccases (SLAC) (Figure 7B) consist of two, laccases and ascorbate oxidases (Figure 7A) of three and ceruloplasmin (Figure 7C) of six homologous domains (Nakamura *et al.* 2003). Plant and fungal laccases, like the very first identified laccase from the sap of the Japanese lacquer tree (Yoshida 1883) or the highly active commercially available fungal laccase from *Trametes versicolor*, typically contain three different domains. (Bertrand *et al.* 2002).

The same is valid for ascorbate oxidases, enzymes currently only identified in plants and fungi (Peng *et al.* 2015).

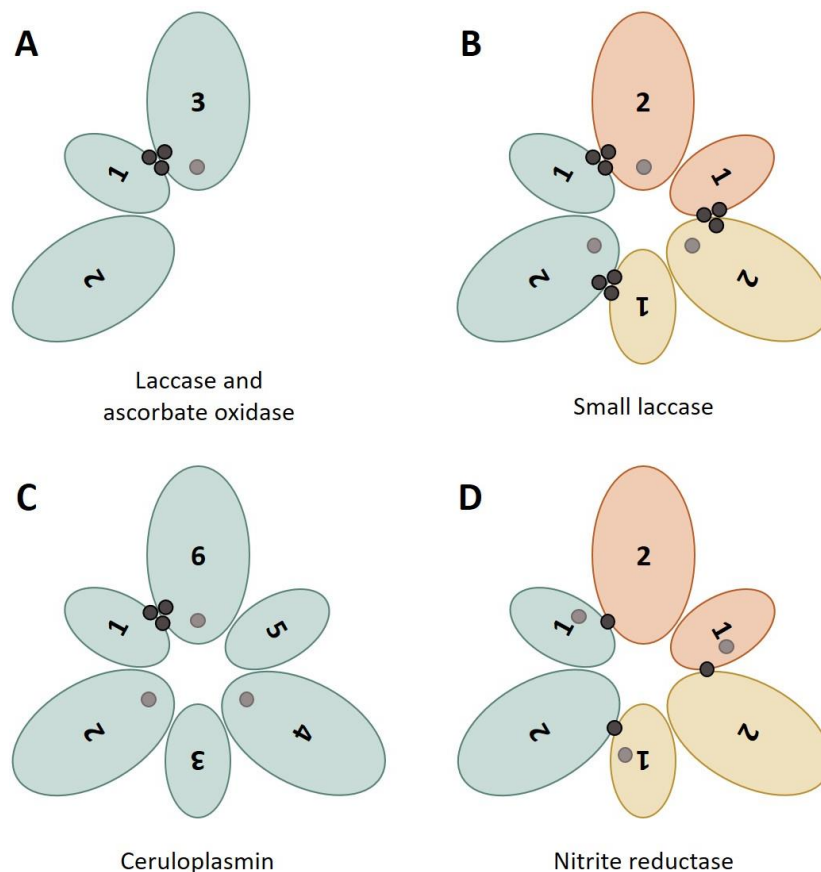


Figure 7: Domain and copper ion organization of several copper-containing enzymes. Small laccases (SLAC) (**B**) and nitrite reductases each form homotrimers by combining three of their two-domain proteins to one active hexagram-like enzyme structure. The type 2 (T2) and type 3 (T3) copper ions are located between the monomeric chains whereas the three type 1 copper ions in the mononuclear copper center (MNC) are each stationed inside one of the two different domains. While SLAC contain one T2 and two T3 coppers in their trinuclear copper center (TNC), nitrite reductases (**D**) only have one T2 copper ion as equivalent to it. Ceruloplasmin (**C**) is shaped like the homotrimer of SLAC except for being composed of one large protein chain containing only one single TNC between the first and last domain instead of three individual ones. Laccases and ascorbate oxidases (**A**) contain three of these domains with the TNC in-between domains one and three and the MNC integrated in domain three. Monomers are distinguished by different colors while copper ions are shown as grey (type 1) and black dots (type 2 and 3). Schematics were adapted from literature (Skálová *et al.* 2009; Mot & Silaghi-Dumitrescu 2012).

Small laccases (SLAC) are two-domain MCOs (2dMCOs) that shape a homotrimer as their catalytically active form by combining three protein chains in a hexagram-like configuration. The T2 and T3 copper ions are located between the monomeric chains whereas the three mononuclear copper ions (MNC) are positioned inside one of the two different domains, respectively (Nakamura *et al.* 2003; Machczynski *et al.* 2004). While SLAC contain one T2 and two T3 coppers in their trinuclear copper center (TNC), the structurally homologous non-MCO nitrite reductases only have one T2 copper ion as equivalent to it (Skálová *et al.* 2009).

Ceruloplasmin possesses a similar structure as the homotrimer of SLAC, but with only one single TNC between the first and last domain of its six-domain protein chain. Laccases and ascorbate oxidases count as three-domain MCOs (3dMCO) with the TNC between domain 1 (D1) and 3 (D3) and the MNC incorporated into D3 (Morozova *et al.* 2007; Skálová *et al.* 2009; Mot & Silaghi-Dumitrescu 2012). Domain 2 (D2) connects D1 and D3 while partially building the access channel to the TNC, where the electrons from T1 are successively directed to (Tepper *et al.* 2011).

These domains from all types of MCO are called cupredoxin-like. Cupredoxins form a distinctive β -sandwich structure that consists of 6 to 13 β -strands arranged into an antiparallel Greek-key motif (Figure 8A) (Messerschmidt 2010).

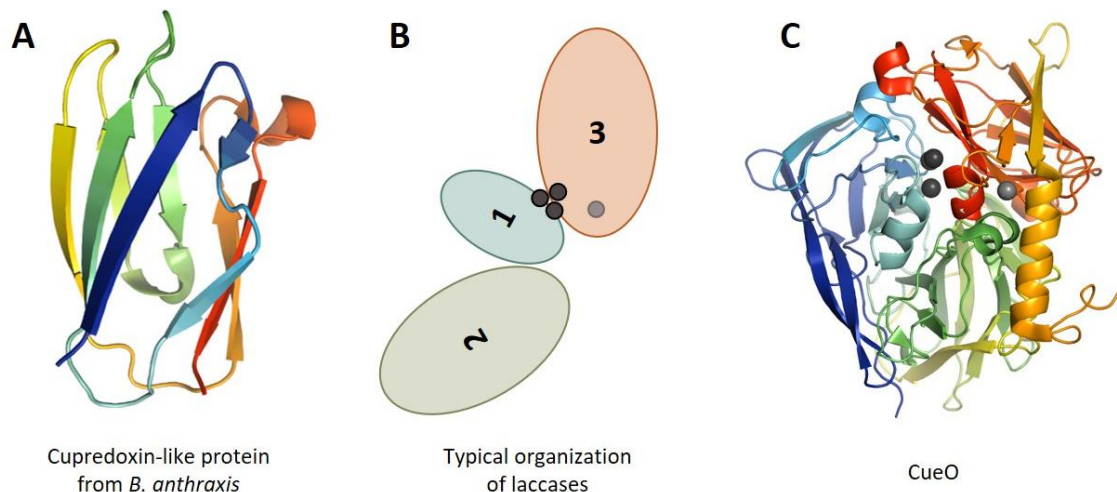


Figure 8: Typical structure of laccases and their cupredoxin-like domains with the example of CueO from *E. coli*. The eight β -strands of the cupredoxin-like protein from *Bacillus anthracis* (A) form a β -sandwich shape arranged into a Greek-key motif (PDB ID: 4HCF) (Kim *et al.* 2012) which is the typical composition of the three laccase domains (B) (Messerschmidt 2010). The ribbon diagram of CueO from *E. coli* (C) (PDB ID: 3PAV) (Montfort *et al.* 2011) shows the arrangement of the three cupredoxin-like domains with the Trinuclear copper center (TNC) (black dots) between D1 (blue) and D3 (red), the Mononuclear copper center (MNC) (grey dot) positioned inside D3 while D2 (green) is located slightly in the background. The protein structures were visualized with PyMOL (The PyMOL Molecular Graphics System, Version 2.3.3 Schrödinger, LLC.) in rainbow color scheme.

They are widespread copper-binding proteins linked to electron transfer pathways and naturally possess a large variety of different copper center geometries (Roger *et al.* 2014). The structure of the best studied bacterial MCO at present – CueO from *E. coli* – is similar to those of fungal laccases (e.g. from *Trametes versicolor*) with the addition of a 42-residue insert in D3. This protein section includes a methionine-rich helix that is displayed in yellow in the CueO structure shown in Figure 8C and covers the entrance to the MNC (Roberts *et al.* 2002).

1.4.2 Occurrence and function

MCOs are ubiquitous and modify a broad spectrum of natural and synthetic compounds. Since the ability to use specific substrates overlaps greatly between the different superfamilies, a sole substrate-based distribution into laccases or other subgroups is highly difficult (Reiss *et al.* 2013).

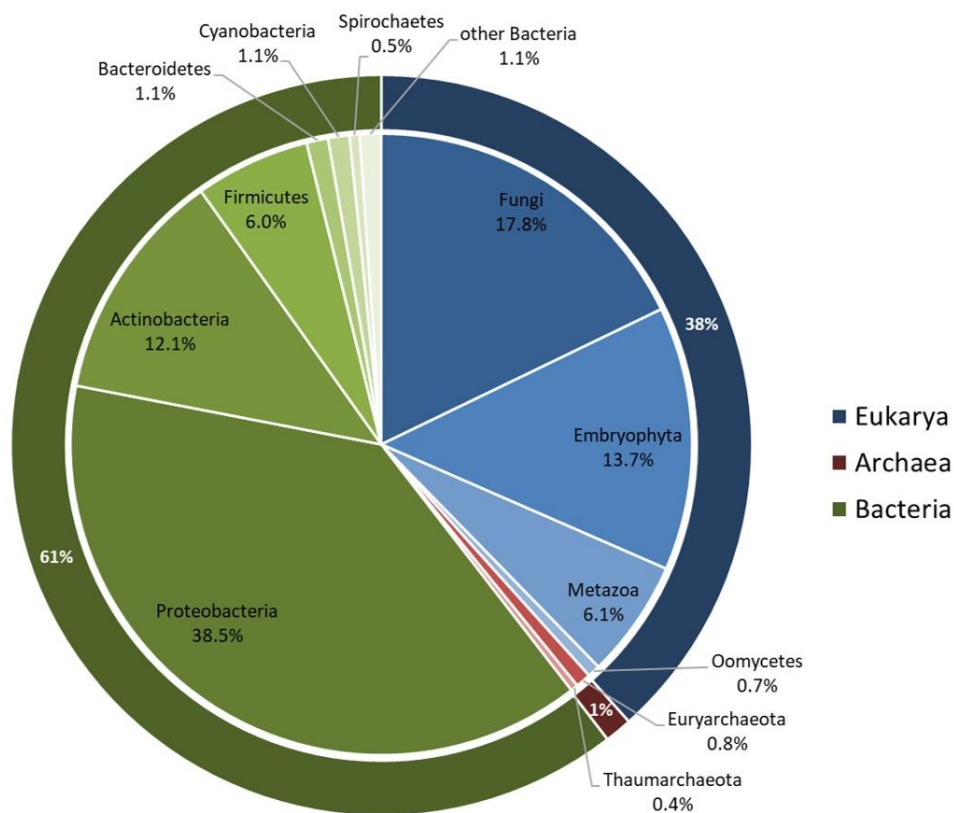


Figure 9: Taxonomic distribution of the MCO sequences in LccED. Taxa with less than 50 sequences (<0,4 %) were disregarded (biocatnet LccED v6.4: <https://lcced.biocatnet.de/> access 2019-07-21; (Sirim *et al.* 2011; Buchholz *et al.* 2016).

Furthermore, MCOs occurred in all three domains, as shown by the taxonomic distribution of their sequences from LccED in Figure 9.

Due to their easier production and purification, however most of the presently characterized laccases were of fungal or bacterial origin (Yin *et al.* 2019). Eukaryotic organisms delivered 38 % of all MCO sequences collected in LccED. Nearly half of these derived from fungi, mainly Ascomycota and Basidiomycota, the other half originated from Embryophyta (land plants), Metazoa and Oomycetes. Archaea only contributed 1 % of the sequences with Euryarcheota and Thaumarchaeota being the main sources. The largest number of sequences was found in the domain of Bacteria. Proteobacteria accounted for most of sequences with 38.5 % while Actinobacteria and Firmicutes follow with 12.1 and 6 %, respectively. Bacteroidetes, Cyanobacteria, Spirochaetes and other bacteria still contributed up to 1.1 % of MCO sequences (biocatnet LccED v6.4: <https://lcced.biocatnet.de/> access 2019-07-21) (Sirim *et al.* 2011; Buchholz *et al.* 2016).

MCOs play a significant role in a lot of metabolic processes like iron and copper homeostasis. Additionally, to representing a main antioxidant in mammals, ceruloplasmin acts as copper depot by carrying the majority of copper ions in the blood. It has the ability to bind metal ions in chelate complexes and functions as an essential iron transport protein in the plasma (Roberts *et al.* 2002; Orzheshkovskiy & Trishchynska 2019).

Until now, SF#6 – ascorbate oxidases – has only been identified in plants and fungi. They have an impact on various signaling pathways by oxidizing ascorbate to monodehydroascorbate and therefore changing the redox state of the plant cells (Hoegger *et al.* 2006; Peng *et al.* 2015).

Laccases or laccase-like MCOs were previously detected in plants, insects, fungi and bacteria, where they have a variety of functions and convert a broad range of substrates like phenols and aromatic amines (Pardo *et al.* 2018; Wang, F *et al.* 2019). In the presence of low-molecular weight compounds that act as electron carriers between enzyme and substrate, laccases are even able to modify substrates that have a higher redox potential or that are normally too large for their active site. Naturally occurring small-sized phenolic mediators and synthetically produced ones, like 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) or 1-Hydroxybenzotriazole, were successfully used for bioremediation and to degrade and demethylate the lignin included in kraft pulp (Cañas & Camarero 2010; Christopher *et al.*

2014). Fungal redox mediators are even capable of enhancing laccase activity to a level that enables the disruption of β -aryl ether bonds, the most dominant linkage in hardwood lignin (Eggert *et al.* 1996).

Previously, laccase genes were identified in a lot of different plant species. They were observed to be involved in lignin and flavonoid biosynthesis by means of oxidative polymerization (Pourcel *et al.* 2005; He *et al.* 2019; Li *et al.* 2019). These substances play a significant role in plant defense. Lignification increases the stability of the plant cell wall effectively limiting pathogenic ingress while flavonoids can protect against microbial assault and serve as feeding deterrents for herbivores (Hu *et al.* 2018; Wang, Q *et al.* 2019).

Insect laccases are involved in cuticle formation, pigmentation and stabilization by oxidizing catecholamines and therefore mediating cross-linking reactions between proteins and chitin fibers (Asano *et al.* 2019). Laccases from fungal organisms commonly account for larger redox potentials of up to 790 mV. The *Trametes versicolor* laccase being the highest with 785 mV among others from Asco- and Basidiomycetes. Bacterial and plant representatives typically exhibit lower redox potentials of 375-500 mV (Mate & Alcalde 2015; Yin *et al.* 2019). Fungal laccases play a role in morphogenesis, pathogen-host-interaction, stress defense and lignin degradation. They can be found intra- and in large part also extracellular depending on their physiological function (Thurston 1994; Baldrian 2006).

Laccases or laccase-like MCOs from prokaryotic organisms are mostly located intracellularly and involved in cell pigmentation, morphogenesis and iron homeostasis. The spore coat protein A (CotA) from *Bacillus subtilis* is an essential component of the endospore coat and participates in pigment biosynthesis and protection against UV light and hydrogen peroxide (Sharma *et al.* 2007). Another well studied MCO with laccase-like activity is CueO, which is part of the regulatory system for copper homeostasis in *E. coli*. The corresponding gene is located on the *cue* operon which also codes for the copper efflux P-type ATPase CopA. The dysfunction of a CopA homologue in humans is the reason for severe copper disorders (Roberts *et al.* 2002). In addition to its MCO activity, CueO also shows the capability to oxidase Cu^+ to the less toxic Cu^{2+} at another active site located on the methionine-rich helix that covers the entrance to the MNC (Wang *et al.* 2018).

1.4.3 Industrial fields of application

Compared to other oxidative enzymes, laccases and laccase-like MCO possess valuable advantages because they are highly stable and their sole byproduct is water (Reiss *et al.* 2013). Although fungal laccases have been extensively utilized due to their high redox potential, their usability is limited by long fermentation periods, low enzyme yield and restricted reaction conditions. While their bacterial counterparts are less catalytically efficient, they show beneficial characteristics that makes them more interesting for industrial applications. These benefits include a broad substrate specificity and a relatively rapid enzyme production. They are also active at a broad pH and temperature range and show a huge stability in the presence of inhibitory agents (Christopher *et al.* 2014; Chauhan *et al.* 2017).

Due to these properties, laccases show potential in a wide area of application like food, pulp and paper, textile, cosmetics and nanobiotechnology (Figure 10).

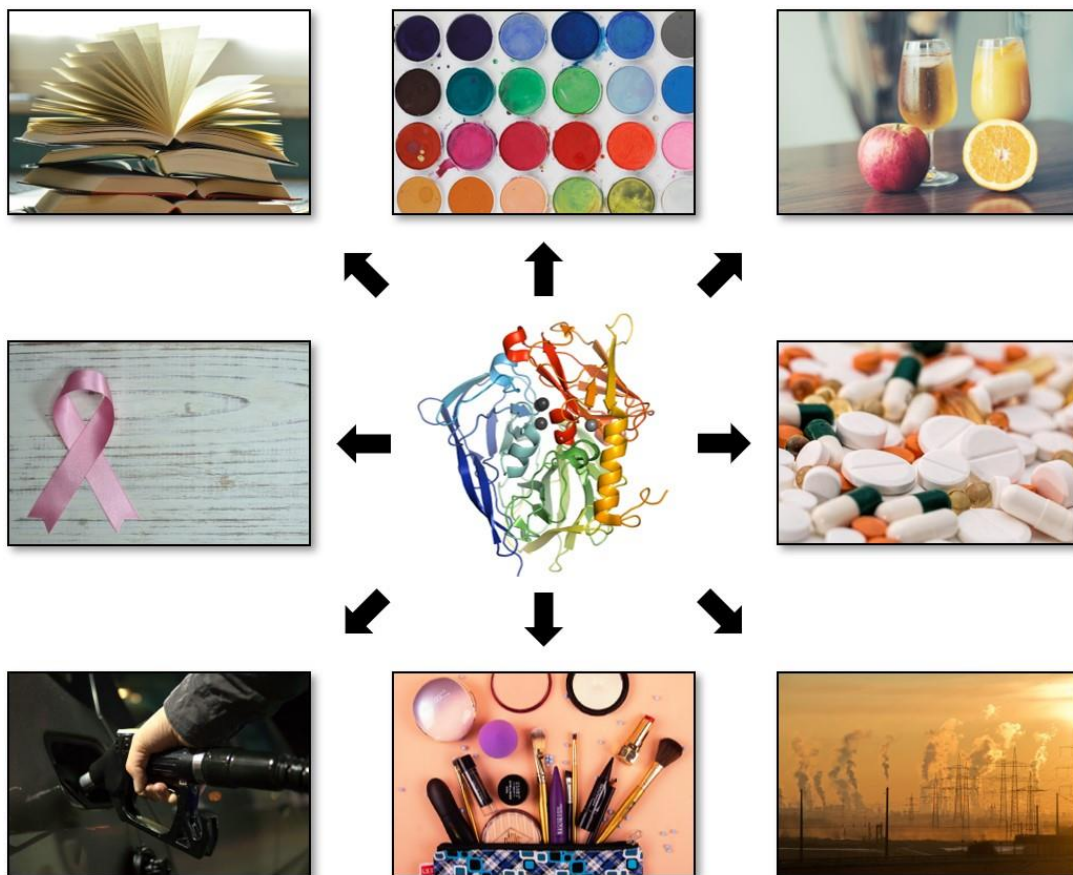


Figure 10: Prospective fields of industrial laccase application. Laccases can potentially be utilized in a variety of industrial applications, such as pulp and paper production, dye decolorization, clarification of juice and wine, biomedicine, bioremediation, cosmetics, biofuels and even cancer treatment (PDB ID: 3PAV (Montfort *et al.* 2011); www.pexels.com (license free)).

For instance, this includes polymer synthesis, bioremediation, clarification of juice and wine as well as dye decolorization and bleaching (Sharma *et al.* 2007; Dubé *et al.* 2008; Zimbardi *et al.* 2016).

Another promising field of application is the production of second-generation biofuels where laccases can be used to remove inhibiting phenolic compounds among other things. They can also be utilized in the biomedical and cosmetic industry for producing diagnostic tests for fungal infections, hair dyes or skin lightening agents (Schneider *et al.* 2012; Mate & Alcalde 2015). Some basidiomycete laccases even show anti-cancerous activity against breast and other carcinoma cells presumably by degrading phenolic hormones involved in malignant cell proliferation (Guest & Rashid 2016).

Laccase activity can be improved considerably by certain beneficial factors. It was shown that the catalytic lifetime of laccases can be prolonged significantly by immobilization (Hublik & Schinner 2000) while their substrate range increases to encompass non-phenolic lignin compounds, activated alcohols like sugar derivatives and even unsaturated lipids by the use of laccase-mediator systems (Witayakran & Ragauskas 2009).

Ultimately, interesting laccases can be genetically engineered with modern scientific tools to be even better suited for industrial applications. Single AAs or whole enzyme domains can be exchanged to enhance properties like heterologous expression, catalytic efficiency and tolerance to unfavorable pH, temperature and other inhibitory conditions (Mate & Alcalde 2015; Pardo *et al.* 2018).

1.5 Aim of the study

As previously introduced, laccases and laccase-like MCO – especially from bacterial origin – indicate a massive metabolic potential with a large variety of possible industrial applications. For instance, lignin represents a significant portion of lignocellulosic biomass that is essentially unutilized at present. Lignolytic enzymes potentially facilitate its efficient exploitation and enable a renewable replacement for fossil-based resources in the future (Christopher *et al.* 2014; Chauhan *et al.* 2017).

Over the course of this thesis, three main goals were intended to achieve (Figure 11).

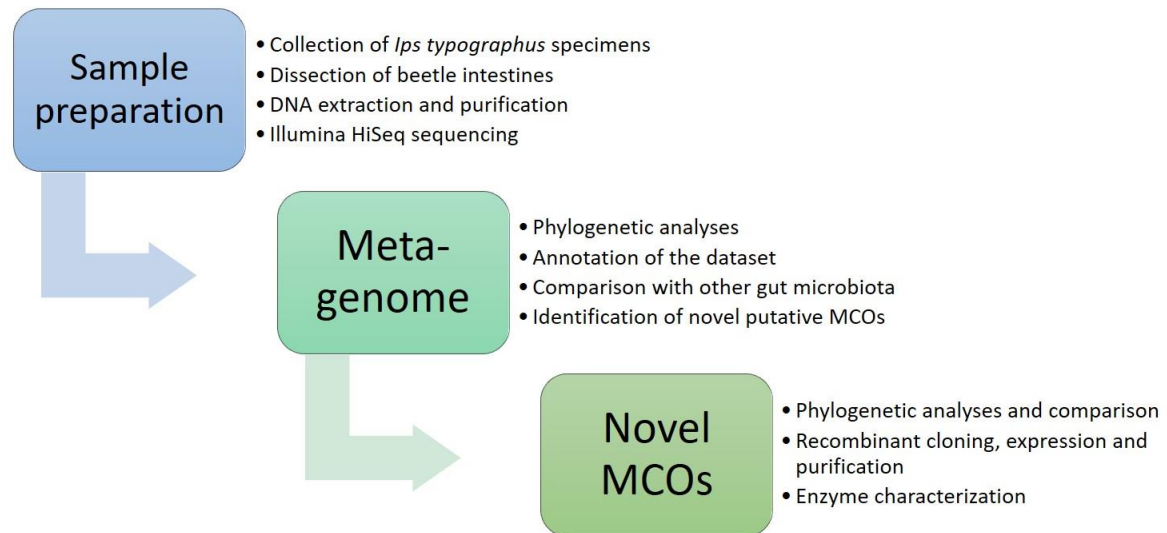


Figure 11: Overview of critical steps this study aimed at. This thesis aimed to accomplish three key steps: The collection and preparation of bark beetle specimen of the species *Ips typographus* including intestinal dissection, DNA extraction and purification as well as subsequent Illumina HiSeq sequencing. The analysis of the resulting metagenome regarding phylogeny prior to annotation and comparison of the dataset with different microbiota. And lastly, the identification, recombinant cloning, heterologous expression, and characterization of novel MCOs from this highly specialized environment.

The first objective was the preparation of the *Ips typographus* samples including collection of specimens, dissection of beetles, extraction, and purification of DNA from the intestines and Illumina HiSeq sequencing of the metagenome.

Next, the resulting metagenomic dataset was planned to be analyzed phylogenetically, followed by annotation and storage within the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012). After the comparison with other gut-associated microbiota on phylogenetic and functional level, yet unknown putative MCO genes were intended to be discovered from the *Ips typographus* microbiome and several bacterial datasets *in silico*. These novel enzymes were then meant to be analyzed and compared to each other as well as to already known MCOs prior to recombinant cloning, expression, and purification of the proteins. Finally, a characterization of these enzymes regarding structure, storage behavior, temperature and pH preference was planned.

2 Material und Methods

2.1 Bark beetle sample collection and dissection

The bark beetle specimens used in this study were collected after luring them into Theysohn slot traps (Niemeyer *et al.* 1983) by means of the commercially available European spruce bark beetle pheromone Ipsowit® (Witasek, Feldkirchen, Austria) (Figure 12). After collection they were immediately stored at -20 °C.



Figure 12: The type of slot trap and bark beetle pheromone used in this study. The *Ips typographus* pheromone Ipsowit® (A) (Witasek, Feldkirchen, Austria) was positioned inside the slot trap (B). Due to the design of the trap, the bark beetles cannot leave the trap after entering through the slots. Captured specimens can be collected from the compartment at the bottom of the trap (Flechtmann *et al.* 2000).

Before dissection, the exterior of the beetles was disinfected with ethanol and they were fixated on paraffin plates with their ventral abdomen up. For the dissection, the immobilized beetles were submerged in sterile phosphate buffered saline (PBS). After opening the abdomen with small tweezers, the intestines were pulled out. Shell pieces were removed before they were collected in 1.5 ml sample tubes and stored at -20 °C.

2.2 Preparation of buffers with specific pH

Citric acid (0.1 M) and disodium phosphate solution (0.2 M) were used to prepare citrate-phosphate buffers between 2.5 and 3.2 pH, according to the following table:

Table 2: Preparation of citrate-phosphate buffer with specific pH

Desired pH	Citric acid 0.1 M [ml]	Disodium phosphate 0.2 M [ml]	H ₂ O [ml]
2.5	45.8	4.2	50
2.8	42.2	7.8	50
3.2	37.7	12.3	50

Table 3: Preparation of acetate buffer 0.1 M with specific pH

Desired pH	Acetic acid 0.2 M [ml]	Sodium acetate 0.2 M [ml]	H ₂ O [ml]
3.6	46.3	3.7	50
4.0	41.0	9.0	50
4.4	30.5	19.5	50
4.8	20.0	30.0	50
5.2	10.5	39.5	50
5.6	4.8	45.2	50

Acetate buffers of varying pH used in this study were prepared by blending a 0.2 M acetic acid solution with a 0.2 M sodium acetate solution and adding the same volume of water. The desired pH of the 0.1 M acetate buffer can be adjusted by changing the ratio of these solutions (Table 3).

A 0.1 M sodium phosphate buffer of specific pH can be prepared with the same method by blending 0.2 M monosodium phosphate with 0.2 M disodium phosphate and adding an equal volume of water (Table 4).

Table 4: Preparation of sodium phosphate buffer 0.1 M with specific pH

Desired pH	Monosodium phosphate 0.2 M [ml]	Disodium phosphate 0.2 M [ml]	H ₂ O [ml]
6.0	87.7	12.3	100
6.4	73.5	26.5	100
6.8	51.0	49.0	100
7.2	28.0	72.0	100
7.6	13.0	87.0	100
8.0	5.3	94.7	100

2.3 Bacterial strains, vectors and primers

The bacterial strains, plasmids and primers used in this thesis are shown in Table 5 to Table 8, respectively.

Table 5: Bacterial strains used in this thesis

Strain	Characteristics	Source/Reference
<i>E. coli</i> DH5 α	<i>supE44</i> , Δ <i>lacU169</i> (Φ 80 <i>lacZ</i> Δ M15) <i>hsdR17 recA1 endA1 gyrA96 thi-1 relA1</i>	Invitrogen (Karlsruhe, Germany)
<i>E. coli</i> BL 21 (DE3)	F' <i>ompT hsdS_B</i> ($r_B^- m_B^-$) <i>gal dcm</i> , (DE3)	Novagen (Darmstadt, Germany)
<i>E. coli</i> T7 SHuffle	F' <i>lac, pro, lacI^q</i> / Δ (<i>ara-leu</i>)7697 <i>araD139 thuA2 lacZ::T7 gene1</i> Δ (<i>phoA</i>) <i>Pvull phoR ahpC* galE</i> (or U) <i>galK</i> λ <i>att::pNEB3-r1-cDsbC</i> (Spec ^R , <i>lacI^q</i>) <i>ΔtrxB rpsL150(Str^R) Δgor Δ(malF)3</i>	New England BioLabs (Frankfurt am Main, Germany)
<i>Duganella</i> sp. HH01	Wild type isolate, amp ^R , tet ^R , gen ^S , kan ^S	(Hornung <i>et al.</i> 2013)
<i>Stenotrophomonas maltophilia</i> K279a	Wild type isolate	(Avison <i>et al.</i> 2000)
<i>Sinorhizobium fredii</i> NGR234	Wild type isolate	(Trinick 1980)

Table 6: Vectors used in this thesis

Vector	Characteristics	Source/Reference
pET22b(+)	T7 promoter, His-tag, MCS (<i>AvaI</i> , <i>XhoI</i> , <i>NotI</i> , <i>EagI</i> , <i>HindIII</i> , <i>SalI</i> , <i>SacI</i> , <i>EcoRI</i> , <i>BamHI</i> , <i>NcoI</i>), <i>peI_B</i> , <i>lacI</i> , amp ^R	Novagen (Darmstadt, Germany)
pDrive	Cloning vector, F', <i>lacZ</i> Δ M15, <i>recA1</i> , <i>lac</i> , <i>thi-1</i> , <i>relA1</i> , Amp ^R	QIAGEN (Hilden, Germany)

Table 7: Plasmids used and constructed in this thesis

Plasmid	Description	Source/Reference
pET22b::StrepII	pET22b(+) with StrepII-tag in MCS, stop codon upstream of His-tag	Evocatal (Monheim am Rhein, Germany)
pET22b:: <i>cueo</i> ::StrepII	Laccase CueO from <i>E. coli</i> in pET22b::StrepII	Evocatal (Monheim am Rhein, Germany)
pET22b-SDM::StrepII	Modified pET22b::StrepII; <i>NheI</i> recognition site inserted 5' of <i>NdeI</i> by site-directed mutagenesis	This study
pEX-K4::1282	Potential MCO <i>lps1282</i> from <i>lps typographus</i> metagenome in pEX-K4; synthesized	Purchased from Eurofins Genomics (Ebersberg, Germany)
pEX-K4::2204	Potential MCO <i>lps2204</i> from <i>lps typographus</i> metagenome in pEX-K4; synthesized	Purchased from Eurofins Genomics (Ebersberg, Germany)
pEX-K4::28714	Potential MCO <i>lps28714</i> from <i>lps typographus</i> metagenome in pEX-K4; synthesized	Purchased from Eurofins Genomics (Ebersberg, Germany)

Plasmid	Description	Source/Reference
pET22b::hh01-218::StrepII	Potential MCO HH01-218 from <i>Duganella</i> sp. HH01 in pET22b::StrepII	This study
pET22b::hh01-482::StrepII	Potential MCO HH01-482 from <i>Duganella</i> sp. HH01 in pET22b::StrepII	This study
pET22b::sm3532::StrepII	Potential MCO Sm3532 from <i>Stenotrophomonas maltophilia</i> in pET22b::StrepII	This study
pET22b::ngr578::StrepII	Potential MCO NGR578 from <i>Sinorhizobium fredii</i> NGR234 in pET22b::StrepII	This study
pET22b-SDM::ngr600::StrepII	Potential MCO NGR600 from <i>Sinorhizobium fredii</i> NGR234 in pET22b-SDM::StrepII	This study
pET22b-SDM::ips1282::StrepII	Potential MCO Ips1282 from <i>Ips typographus</i> metagenome in pET22b-SDM::StrepII	This study
pET22b-SDM::ips2204::StrepII	Potential MCO Ips2204 from <i>Ips typographus</i> metagenome in pET22b-SDM::StrepII	This study
pET22b::ips14138::StrepII	Potential MCO Ips14138 from <i>Ips typographus</i> metagenome in pET22b::StrepII	This study
pET22b-SDM::ips21622::StrepII	Potential MCO Ips21622 from <i>Ips typographus</i> metagenome in pET22b-SDM::StrepII	This study
pET22b-SDM::ips24328::StrepII	Potential MCO Ips24328 from <i>Ips typographus</i> metagenome in pET22b-SDM::StrepII	This study

Table 8: Primer used in this thesis

Primer	Sequence 5' → 3'	GC content [%]	Length [nt]
M13-20_for	GTAAAACGACGGCCAGT	52.94	17
M13_rev	CAGGAAACAGCTATGAC	47.06	17
T7_promoter	TAATACGACTCACTATAGGG	40.00	20
T7_terminator	GCTAGTTATTGCTCAGCGG	52.63	19
pET22b_SDM_for	AGCAGCGGTCGGCAGCAGGCTAGCCATATGTA TATCTCC	56.41	39
pET22b_SDM_rev	CTGCTGCCGACCGCTGCTGCTGGTCTGCTGCT C	69.70	33
HH01-218_Ndel_for	ACACCATATGAATAGCGGGGTAAAAGAGTTCC ACTTG	43.24	37
HH01-218_SacI_rev	ACGAGCTCGTGTCCGCCATGTCCCGACGAT	63.33	30
HH01-482_Ndel_for	ACACCATATGATTTACAGTAGAGACTTCTTCA AGG	40.00	35
HH01-482_SacI_rev	GATCGAGCTCGTGGCCGCTATGGCCGCT	67.86	28
Sm-3532_Ndel_for	ACACCATATGTCGCATGATGATTTTCGTGGTC CACGCG	50.00	38

Primer	Sequence 5' → 3'	GC content [%]	Length [nt]
Sm-3532_SacI_rev	TTGAGAGCTCTTCGTGACCCCTCACCGTGCGCAT	58.82	34
Sm-4696_NdeI_for	ACACCATATGAATACCCGCAATCTCCCCGGTCG	55.88	34
Sm-4696_SacI_rev	GAGCGAGCTCTTCTTCGATCCTCACTTCGCGCATCATGC	56.41	39
NGR147_NdeI_for	ACACCATATGACGCCGTTGCTCAGCCGC	60.71	28
NGR147_HindIII_rev	ACAAGCTTGATAATTCTGAAATAGCCGGTCATTCCG	41.67	36
NGR578_NdeI_for	ACACCATATGTTCAACAGAAGACAGTTGCTCGGC	47.06	34
NGR578_SacI_rev	ACGAGCTCGCCGTGGCTTGAGTGTTCG	60.71	28
NGR600_NheI_for	ACACGCTAGCATGCCAGACTTCTCACGTCGTCAGT	54.29	35
NGR600_SacI_rev	ACGAGCTCGGTTTGATACTCGACCGCCGTCA	58.06	31
NGR688_NdeI_for	ACACCATATGTTCAACAGAAGACAGATACTGGGAGCAG	44.74	38
NGR688_XhoI_rev	ACCTCGAGGCCGTGATTTCCATGTTTCGGCGT	56.25	32
lps1282_NheI_for	AAGGGCTAGCATGTTACGTCGCGAT	52.00	25
lps1282_SacI_rev	TTTTGAGCTCACTCACTGTAAAGCCGAGCA	46.67	30
lps2204_NheI_for	GGATGCTAGCATGCAACGTCGTG	56.52	23
lps2204_SacI_rev	TTTCGAGCTCAGCGCTCACCGTGAAC	57.69	26
lps14138_NdeI_for	ACTGCATATGCATCGTCGTGATTTTCTG	42.86	28
lps14138_SacI_rev	GCCTGAGCTCTGCCACGGTGAAGC	66.67	24
lps21622_NheI_for	GAAGGCTAGCATGCAACGACGCGATTTTTTTG	48.39	31
lps21622_SacI_rev	TTTAGAGCTCGGCCGTGACAGTAAAGCCCG	56.67	30
lps24328_NheI_for	GGACGCTAGCATGAACCGTCGTGATTTTC	53.57	28
lps24328_SacI_rev	AAACGAGCTCCGCCGATACCGTAAATCCAAGC	53.13	32
lps28714_NdeI_for	ACTGCATATGAACCGTCGCGATTTCTGTG	50.00	28
lps28714_SacI_rev	AAGAGAGCTCTGCCGACACCGTAAACGC	57.14	28
B-V3-for-MID1	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGTTACGGRAGGCAGCAG	60.00	55
B-V6-rev-MID1	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCRCACGAGCTGACGAC	59.57	47
E-for-MID3	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACGCACTCGTGCCAGCMGCCGCGGTAA	62.71	59
E-rev-MID3	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGGTGCCCTTCCGTCA	62.22	45
A-for-MID2	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACACCCTAYGGGGYGCASCAG	60.34	58
A-rev-MID2	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGTGCTCCCCCGCCAATTCCT	62.00	50

All primer used in this thesis where ordered from Eurofins MWG Operon (Ebersberg, Germany).

2.4 Cultivation of bacteria

Routinely, *E. coli* strains were cultivated in sterile Luria Bertani (LB) medium with 5 g/l yeast extract, 10 g/l tryptone and 10 g/l sodium chloride (Sambrook & Russell 2001) shaking at 150 to 250 rpm for 18 to 24 h at 37 °C. The exception were preparatory cultures of expression strains carrying vectors with MCO genes. These were grown with the addition of 1 % glucose to block the premature protein expression (Studier 2005). For the use as solid medium, an amount of 1,5 % of agar-agar was added to all media before autoclaving. Bacterial strains were stored for 4-6 weeks at 4 °C on agar plates or mixed 4:1 with glycerol at -70 °C for long-term storage.

2.4.1 Antibiotics and supplements

Depending on purpose and resistance genes on the plasmid, antibiotics and additives were added to the media in appropriate concentrations (Table 9).

Table 9: Used antibiotics, their stock and applied concentrations

Antibiotic	Stock Solution [mg/ml]	Final concentration [µg/ml]	Solvent
Ampicillin (Amp)	100	100	H ₂ O _{bidest}
Kanamycin (Kan)	25	25	H ₂ O _{bidest}

2.4.2 Determination of cell density

The cell density of liquid cultures was measured photometrically (BioPhotometer, Eppendorf, Hamburg, Germany) in one-way cuvettes (10 x 4 x 45 mm, Sarstedt, Nümbrecht, Germany) with 1 cm path length at a wavelength of 600 nm. For *E. coli*, an optical density (OD) of 0.1 corresponds to a cell density of approximately 1×10^8 cells/ml.

2.5 Methods for working with DNA

2.5.1 DNA extraction and purification

In preparation for the DNA extraction, 3-5 ml of the cell culture were harvested by centrifuging for 2 min at 13000 rpm in a tabletop microcentrifuge (minispin Plus, Eppendorf, Hamburg, Germany). The supernatant was discarded, and the DNA was extracted from the pellet by using a DNA isolation kit. The High-Speed Plasmid Mini Kit (Avegene, Taipeh, Taiwan) was utilized to isolate plasmid DNA and Genomic DNA was extracted by using the AquaPure Genomic DNA kit (Bio-Rad Laboratories, Hercules, Canada), both according to manufacturer's instructions.

For the extraction of metagenomic DNA from dissected bark beetle intestines, the tube was first thawed on ice. Then, the DNA was extracted by using the QIAamp[®] DNA Stool Mini Kit (QIAGEN, Hilden, Germany) according to manufacturer's recommendations. The chosen protocol contains lysis conditions to increase the ratio of prokaryotic to eukaryotic DNA.

DNA fragments from PCR or restriction mixtures were purified with the Gel/PCR DNA Fragments Extraction Kit (Avegene, Taipeh, Taiwan) by following the manufacturer's instructions. After measuring the concentration with the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Braunschweig, Germany) the DNA was stored at -20 °C until further use.

2.5.2 Agarose gel electrophoresis

The size of DNA fragments can be determined by using agarose gel electrophoresis and comparing the band of the sample with the specific band pattern of a standardized size marker (GeneRuler[™] 1 kb, Fermentas, St. Leon-Rot, Germany). This method also allows the purification and separation of DNA fragments obtained by PCR, restriction, or DNA isolation. Required solutions are shown in Table 10 and Table 11.

For the gel preparation, 0.8 % of agarose was dissolved in 1x TAE buffer and heated until the solution was clear. For relatively small fragments, a concentration of up to 2 % of agarose was used.

Table 10: TAE buffer (50x)

Components	Quantity
Tris	2 M
EDTA (pH 8.0)	100 mM
H ₂ O _{bidest}	<i>ad</i> 1000 ml
Acetic acid	up to pH 8.1

Table 11: Loading dye

Components	Quantity
Glycerol (30 %)	60 ml
EDTA (pH 8.0)	50 mM
Bromophenol blue	0.5 g
Xylene cyanol	0.5 g
H ₂ O _{bidest}	<i>ad</i> 200 ml

The gel solution was then filled into a cast to cool down and a comb was added to create wells for the samples. After the complete hardening of the gel, it was placed into an electrophoresis chamber (Hoefer™ HE-33 mini horizontal submarine unit, Amersham Biosciences, San Francisco, USA) filled with 1x TAE buffer.

The samples were mixed with loading dye and filled into the wells. A power supply (EPS 301, Amersham Biosciences, San Francisco, USA) was used to perform the electrophoresis at 90 – 120 V for 22 – 90 min depending on fragment size and further purpose of the DNA. Afterwards, the gels were stained for 5 – 15 min in an ethidium bromide solution (~10 µg/ml) and destained in water to remove the excess. For visualization with UV light, the gel documentation unit Molecular Imager® Gel Doc™ XR+ (Bio-Rad Laboratories, Munich Germany) was used. Documentation was performed with the software Quantity One 4.6.9 (Bio-Rad Laboratories, Munich Germany).

If the DNA fragment is designated for further use, it can then be cut out of the gel and isolated with the Gel/PCR DNA Fragments Extraction Kit from Axygen (Taipei, Taiwan). In this case it is recommended to shorten the staining time in ethidium bromide and the exposure to UV light to lessen the damage to the DNA.

2.5.3 Polymerase chain reaction (PCR)


The method of polymerase chain reaction (PCR) was used to amplify the MCO gene sequences in preparation for ligation or to test possible clones for the intended insert gene. The PCR mixture was prepared by using a standard scheme (Table 12) and

was then processed in a thermocycler (Mastercycler Personal, Eppendorf, Hamburg, Germany) with the temperature steps shown in Table 13. The amplified DNA fragments were made visible by using agarose gel electrophoresis (cf. 2.5.2).

Table 12: Standard recipe for 100 μ l of PCR mix

Components	Volume [μ l]
<i>taq</i> buffer 10x	10
dNTPs (10 mM)	2
Primer forward (10 μ M)	2
Primer reverse (10 μ M)	2
<i>taq</i> polymerase	2
Template DNA	1
H ₂ O _{bidest}	<i>ad</i> 100

Table 13: Standard PCR program

Reaction steps	Temperature [°C]	Duration [min:s]	Cycles
Initial denaturation	95	5:00	
Denaturation	95	1:00	
Annealing	█	0:30	
Elongation	72	█	
Final elongation	72	5:00	


Annealing temperature and elongation time vary depending on primer type and amplicon size.

The annealing temperature (T_A) usually lies about 5 °C under the melting temperature (T_M) of the primer pair. When the forward and reverse primers don't have the same melting temperature, the lower one was used. The elongation time depends on the reaction time of the polymerase and the size of the amplicon. While the *taq* polymerase converts approximately 1 kb every 60 seconds, the Phusion polymerase works much faster with 15-30 s/kb depending on template configuration. In contrast to the first mentioned, the Phusion also exhibits a proof-reading function and was therefore primarily used to amplify genes prior to cloning or for 16S and 18S rRNA sequencing. Reaction mixture (Table 14) and PCR program (Table 15) of the Phusion DNA polymerase differ from the *taq* polymerase, which was mainly utilized for checking the size of inserts or in general when no proof reading was needed.

Table 14: Standard recipe for 50 µl of Phusion PCR mix

Components	Volume [µl]
5x Phusion buffer	10
dNTPs (10 mM)	1
Primer forward (10 µM)	2.5
Primer reverse (10 µM)	2.5
Phusion DNA polymerase	2
Template DNA	1
H ₂ O _{bidest}	<i>ad 50</i>

Table 15: Phusion PCR program

Reaction steps	Temperature [°C]	Duration [min:s]	Cycles
Initial denaturation	98	0:30	
Denaturation	98	0:10	
Annealing	█	0:30	
Elongation	72	█	
Final elongation	72	5:00	

Annealing temperature and elongation time vary depending on primer type and amplicon size.

Direct colony PCR

For checking several clones at the same time after transformation, a direct colony PCR was performed. For this purpose, the PCR mix was prepared with a primer pair that binds to the used plasmid and was then evenly divided into several reaction tubes. Instead of template DNA, a sample of colony material was added directly into the tube with a sterile toothpick. This mix was then processed as mentioned above. To secure the potentially significant clones, they were also plated on LB agar plates, incubated, and temporarily stored at 4 °C.

2.5.4 Digestion of DNA by restriction endonucleases

Type II restriction endonucleases were used to linearize plasmids or to cut out their inserts at specific sites recognized by the used enzyme. For this purpose, a reaction mixture was prepared with a volume of 20 µl for analytical and 50 µl for preparative

application (Table 16). After incubation at 37 °C for 1 - 3 h depending on the enzyme, the cleavage was tested by agarose gel electrophoresis. All endonucleases and their buffers used in this study were supplied by Fermentas (St. Leon-Rot, Germany) or New England BioLabs (Frankfurt am Main, Germany).

Table 16: Analytical and preparative reaction mixture for the digestion of DNA by restriction endonucleases

Components	Analytical Reaction [μl]	Preparative Reaction [μl]
Restriction endonuclease	0.5	1
Enzyme buffer 10x	2	5
DNA	5	15
H ₂ O _{bidest}	<i>ad 20</i>	<i>ad 50</i>

For small amounts of DNA, it was possible to digested with two different restriction endonucleases in the same reaction. Suitable buffers and enzyme ratios for the individual combinations were selected according to manufacturer's recommendations.

2.5.5 Ligation of DNA

An insert can be integrated into a correspondingly cleaved plasmid by use of T4 DNA ligase (Fermentas, St. Leon-Rot, Germany). A fragment that was digested with the same endonuclease as the plasmid can be ligated directly by using a standard mixture (Table 17) at 4 °C overnight or at 16 °C for 3 h. Vector and insert DNA was added in a 1:3 to 1:6 ratio depending on insert size.

Table 17: Standard ligation reaction

Components	Volume [μl]
T4 DNA Ligase Buffer 10x	2
T4 DNA Ligase	0.5
Plasmid DNA	x
Insert DNA	y
H ₂ O _{bidest}	<i>ad 20</i>

Purified DNA fragments that resulted from *taq* PCR were subcloned into the pDrive cloning vector using the PCR Cloning Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions. The reaction mixture (Table 18) was incubated at 16 °C for 2 h and subsequently transformed into chemically competent cells.

Table 18: Composition of a pDrive ligation mixture

Components	Volume [μ l]
pDrive cloning vector	0.5
PCR product	0.5-2
Ligation master mix	2.5
H ₂ O _{bidest}	<i>ad 5</i>

Because of glucose in the LB medium, the blue-white screening function of the pDrive cloning vector was ineffective. The addition of glucose blocks the expression of the β -galactosidase gene on the plasmid. Since this protein is responsible for the blue coloring, all colonies remain white whether they carry an inserted gene or not.

2.5.6 Preparation of chemically competent cells

To create *E. coli* cells susceptible for heat shock transformation, they were treated with rubidium chloride. Buffers TFB1 and TFB2 were prepared (Table 19 and Table 20), filtered sterile and stored at 4 °C. A preparatory culture of the target strain was used to inoculate 250 ml of LB broth, which was followed by shaking incubation at 37 °C. After reaching an OD₆₀₀ of 0.5, the cells were harvested by centrifugation at 5000 rpm and 4 °C for 20 min (Sorvall RC6+, rotor F10S-6x500y, Thermo Fisher Scientific, Braunschweig, Germany) and resuspended in 75 ml of the TFB1 buffer. Next, the cells were cooled for 90 min on ice and again centrifuged before the current buffer was exchanged with 10 ml of TFB2. In aliquots of 100 μ l, this mixture was filled into pre-cooled 1.5 ml reaction tubes and stored at -70 °C.

Table 19: TFB1 buffer

Components	Quantity
RbCl	100 mM
MnCl ₂ * 4 H ₂ O	50 mM
KAc	30 mM
CaCl ₂ * 2 H ₂ O	10 mM
Glycerol	15 %
Acetic acid (1 %)	to pH 5.8
H ₂ O _{bidest}	ad 75 ml

Table 20: TFB2 buffer

Components	Quantity
MOPS	10 mM
RbCl	10 mM
CaCl ₂ * 2 H ₂ O	75 mM
Glycerol	15 %
KOH (10 mM)	to pH 6.8
H ₂ O _{bidest}	ad 20 ml

2.5.7 Heat-shock transformation

After ligation of vector and insert DNA, the plasmid was transferred into chemically competent cells. The *E. coli* strain DH5 α was used for cloning purposes. Whereas expression strains like *E. coli* BL21(DE3) or T7 SHuffle were utilized with the aim of protein production.

To prepare for the transformation, the chemically competent cells were put on ice for 5 to 10 min to gently thaw. Then, 1 to 5 μ l of ligation mixture were added and mixed in before incubating them on ice for 30 min. After heat shock treatment for 90 s at 42 °C, the cells were again put on ice for 3 min. Following the addition of 500 μ l of LB broth, the mixture was shaken for 45 min at 37 °C and plated on LB agar plates in aliquots of different volumes. The added antibiotic was chosen depending on the resistance gene on the plasmid. After an over-night incubation at 37 °C the colonies could be picked and tested by direct colony PCR.

2.5.8 Site-directed mutagenesis

With site-directed mutagenesis (SDM), a short DNA sequence of a plasmid can be exchanged or inserted additionally. This method was carried out based on the GeneTailor™ Site-Directed Mutagenesis System (Invitrogen, Karlsruhe, Germany). First, specific primers of approximately 30 nucleotides length were designed (Figure 13). Both contained an overlapping part on the 5' end, which ensured the end-joining

of the amplified DNA fragments, and an extended region to bind to the template plasmid. The mutation site was added to the primer that binds to the target region. To ensure its digestion upon entering the host cell, the plasmid DNA was methylated before amplification. After this step, a PCR reaction was performed with the Phusion DNA polymerase and the plasmid as template. Because the circular plasmid has a multiple times higher transformation rate, this was followed by the digestion of the methylated template DNA with the restriction endonuclease *DpnI*. The PCR product, linear double-stranded DNA containing the mutation, was then transformed into *E. coli* DH5 α and circularized by the hosts cells (Figure 14).

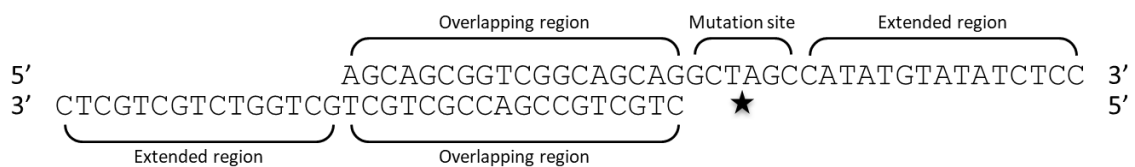


Figure 13: Primer design for site-directed mutagenesis. Both primers contain an overlapping area on the 5' end and an extended region on the 3' end which binds to the target plasmid. The mutation site (marked with the star) only needs to be added to one primer (based on the GeneTailor™ instruction manual, Invitrogen, Karlsruhe, Germany).

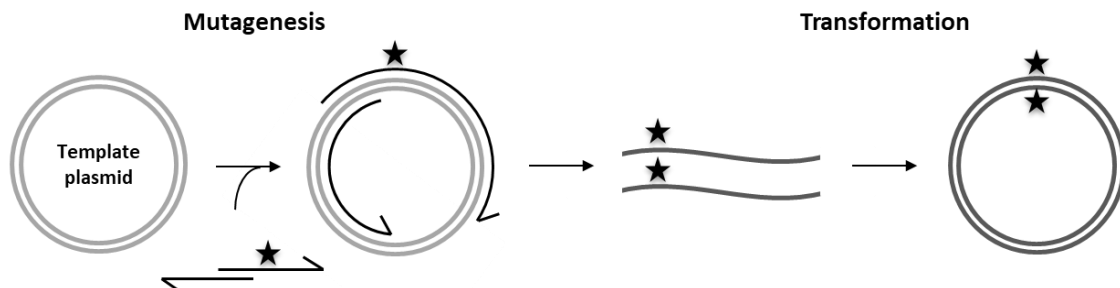


Figure 14: Systematic overview of the site-directed mutagenesis procedure. This method includes the performance of a Phusion PCR reaction with the template plasmid and specifically designed primers that contain an overlapping area, the mutation site (marked with the star), and an extended region for template binding. The product of the mutagenesis is linear double-stranded DNA which contains the mutation site on both strands and is circulated by the host cell upon transformation (based on the GeneTailor™ instruction manual, Invitrogen, Karlsruhe, Germany).

2.6 Metagenomics

2.6.1 Illumina Hiseq sequencing and database analyses

The sequencing of the metagenomic *Ips typographus* DNA was carried out by the Heinrich Pette Institut in Hamburg, Germany at the end of 2014. A genomic library

was created with the isolated bark beetle DNA by using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England BioLabs, Frankfurt am Main, Germany) according to manufacturer's recommendations. Then, the size and quality of the library was checked with a BioAnalyzer High Sensitivity Chip. The 2 x 125 bp v4 HiSeq sequencing run was done with an Illumina HiSeq 2500 instrument in paired end mode. The generated data was processed with the Trimmomatic software (v.0.32) (Bolger *et al.* 2014) to remove the adapter sequences and for a general quality trimming. For the assembly of the 125 bp long reads, the IDBA-UD software (v.1.1.1) (Peng *et al.* 2012) was used.

The taxonomic analysis of the whole metagenomic dataset with MG-RAST was done on 9th December 2014 while the analyses in the Integrated Microbial Genomes (IMG) database were carried out between January 2015 and July 2016.

2.6.2 Phylogenetic analyses

The phylogenetic analyses were carried out by using especially designed primer pairs for the amplification of the bacterial and archaeal 16S rRNA as well as the eucaryal 18S rRNA (cf. chapter 2.3). These primers were composed of Roche 454 pyrosequencing adaptors, key and MID codes that were used to distinguish between the amplicates. Another part of them was a sequence that binds specifically to the rRNA of bacteria, archaea, and eukaryotes, respectively. The PCR reaction was carried out with an annealing temperature of 58.5 °C for the bacterial and 64.5 °C for the eucaryal primers while using a Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Braunschweig, Germany). After purification of the amplicons, the samples were sent to the Göttingen Genomics Laboratory for pyrosequencing. The Roche GS-FLX+454 pyrosequencer and titanium chemistry (Roche, Branford, USA) was used for the 16S and 18S rRNA sequencing.

Obtained sequences were computed by the SILVAngs analysis pipeline (SILVAngs 1.2) for taxonomic classification. During this, the reads were aligned with the SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN, revision 21008) before excluding the ones that didn't meet the standard criteria of length, homopolymerization, ambiguity and alignment quality (Pruesse *et al.* 2012; Quast *et al.* 2013). Also, the reads were dereplicated and unique ones were clustered by using the pipelines software CD-hit (version 3.1.2) (Li & Godzik 2006). The Basic Local Alignment Search Tool

BLASTn (version 2.2.28+) (Camacho *et al.* 2009) was used to classify the operational taxonomic units (OTUs) by comparison to sequences stored in the SILVA database (release 119; <http://www.arb-silva.de>). Reads with values of sequence identity plus coverage divided by 2 below 93 % remained unclassified and were graded as “no relative” in the final assessment of the taxonomic classification by SILVAngs. The sequence processing with SILVAngs was done in February 2015 while BLASTn searches of the resulting 16S and 18S rRNA OTUs were computed in February 2020.

2.7 Protein biochemical methods

2.7.1 Expression using auto-induction medium

Due to the potential toxicity of large protein concentrations or the copper ions necessary during expression, an auto-induction medium was used that blocked unintentional gene expression in the growth phase via 0.05 % of glucose (Table 24) (Studier 2005). The ZY medium (Table 21) was augmented 1/50 with the 5052 solution (Table 23), which raised the concentration of glucose in the medium to 0.05 %, lactose to 0.2 %, and glycerol to 0.5 %. For buffering purposes, 50 mM phosphate was supplemented by adding 1/50th of the M solution (Table 22).

A concentration of 1 mM MgSO₄ promoted a higher cell density in over-night cultures. Depending on the resistance gene on the plasmid, a suitable amount of the corresponding antibiotic was also included. The medium was inoculated 1/200 with a preculture or colony material and incubated shaking at 37 °C to an OD₆₀₀ = 0.6. At this point, CuSO₄ was added to a total concentration of 2 mM. The copper ions were needed during the expression to be integrated into the active center of the protein to promote the right folding.

After reaching an OD₆₀₀ of about 1.0, the glucose of the medium was metabolized, and the lactose started inducing the *lac* promoter. Afterwards, the temperature was changed to 28 °C for slower over-night expression.

Table 21: ZY medium

Components	Quantity
Yeast extract	5 g
Tryptone	10 g
H ₂ O _{bidest}	<i>ad</i> 1000 ml

Table 23: 5052 solution (50x)

Components	Quantity
Glycerol	25 %
Glucose	2.5 %
Lactose	10 %
H ₂ O _{bidest}	<i>ad</i> 50 ml

Table 22: M solution (50x)

Components	Quantity
Na ₂ HPO ₄ * 2H ₂ O	1.25 M
KH ₂ PO ₄	1.25 M
NH ₄ Cl	2.5 M
Na ₂ SO ₄	0.25 M
H ₂ O _{bidest}	<i>ad</i> 50 ml

Table 24: Auto-induction medium

Components	Quantity
5052 solution (50x)	20 ml
M solution (50x)	20 ml
MgSO ₄	1 mM
Ampicillin	100 µg/ml
ZY medium	<i>ad</i> 1000 ml

2.7.2 Crude cell extract preparation

Following the over-night expression of the target genes, the cells were harvested by centrifugation at 5000 rpm and 4 °C for 20 min (Sorvall RC6+, rotor F10S-6x500y, Thermo Fisher Scientific, Braunschweig, Germany). The supernatant was discarded, and the cells were resuspended in phosphate buffered saline (PBS; 20 mM sodium phosphate, 280 mM NaCl, 6 mM potassium chloride, pH 7.4). After the addition of 1 mM phenylmethanesulfonyl fluoride (PMSF) to inhibit protease activity, the cells were immediately disrupted by passing them three times through a French pressure cell press system (American Instrument Company, Silver Spring, Maryland, USA). The lysate was transferred into falcon tubes and centrifuged at 5000 rpm and 4 °C for 10 min to roughly separate the cell debris. The supernatant containing the soluble protein was then filled into 1.5 ml sample tubes and again centrifuged, this time at 13000 rpm and 4 °C for about 1 h until there was no visible clouding left.

2.7.3 Protein purification by Fast Protein Liquid Chromatography

The clear supernatant was suitable for the purification by fast protein liquid chromatography (FPLC). PBS with a pH of 7.4 was used as binding buffer as well, because a neutral pH supports the stability of bacterial MCO enzymes (Chauhan *et al.* 2017). The relatively mild conditions of binding and elution buffers also facilitated the preservation of the intact protein structure. After preparing the required buffers (Table 25 to Table 27) and solutions (H_2O_{bidest} and ethanol 20 %), they were degassed and stored at 4 °C.

Table 25: StrepTrap™ binding buffer (PBS)

Components	Quantity
Na_3PO_4	20 mM
NaCl	280 mM
KCl	6 mM
HCl	to pH 7.4
H_2O_{bidest}	<i>ad</i> 1000 ml

Table 26: StrepTrap™ elution buffer

Components	Quantity
Desthiobiotin	2.5 mM
Binding buffer	<i>ad</i> 500 ml

Table 27: StrepTrap™ regeneration buffer

Components	Quantity
NaOH	0.5 M
H_2O_{bidest}	<i>ad</i> 500 ml

StrepTrap™ HP columns (GE Healthcare, Freiburg, Germany), prepacked with 1 or 5 ml of StrepTactin Sepharose™ High Performance, were stored on 20 % ethanol, and operated with the liquid chromatography system ÄKTAdesign™ (Amersham Biosciences, San Francisco, USA) following the manufacturer's instructions regarding pressure and storage conditions.

The columns were mounted and washed with at least 5 column volumes (CV) of H_2O_{bidest} to remove the storage ethanol. After equilibration with 5 CV of binding buffer, the protein solution was injected and washed off again with 5 to 10 CV binding buffer or until the absorption level was not visibly changing anymore. After elution with 2.5 mM desthiobiotin in PBS, the column was regenerated by alternating between the injection of H_2O_{bidest} , NaOH 0.5 M and H_2O_{bidest} once more.

2.7.4 Protein quantification

The concentration of protein solutions was measured by using the Bradford protein assay (Bradford 1976). This method utilizes the shift in absorption of the Coomassie Brilliant Blue G-250 dye from 465 nm to 595 nm when it builds a complex with protein side chains. To correlate the protein amount with the measured extinction, a reference curve was generated with different concentrations of BSA (bovine serum albumin).

For this assay, 10 μ l of the protein solution were mixed with 490 μ l of the Bradford reagent and incubated for 10 min in the dark. The extinction was measured at a wavelength of 595 nm and PBS was used as reference. Samples with an extinction above the linear range of the reference curve were diluted with PBS.

2.7.5 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by their molecular weight by SDS-Polyacrylamide gel electrophoresis (SDS-PAGE). SDS denatures proteins and binds to their sulfate groups, masquerading their natural charge and turning it negative. Due to this, the proteins can be separated inside an electrical field by size and not by charge. This discontinuous differentiation takes place in a gel with two polyacrylamide concentrations. Contained ammonium persulfate (APS) acts as a linking compound, whereas tetramethylethylenediamine (TEMED) catalyzes the polymerization reaction.

Because the MCOs identified in this study were about the size of 50-60 kDa, resolving gels with 12 % acrylamide were used (Table 28 to Table 30). The components of the resolving gels were mixed and immediately poured between previously cleaned glass plates that were placed in a stand. Isopropyl alcohol was used to level the gel edge and then removed after polymerization. After rinsing with water and decanting, the stacking gel solution was prepared and poured above the resolving gel. A comb was used to form pockets for sample loading. The gel was used immediately or stored at 4 °C for up to 7 days between wet tissues.

Table 28: Standard recipe for a 12 % SDS-PAGE gel

Component	Resolving gel (12 %)	Stacking gel (5 %)
Acrylamide (40 %)	1.5 ml	0.25 ml
Resolving gel buffer 4x	1.25 ml	-
Stacking gel buffer 4x	-	0.50 ml
TEMED	5 μ l	3 μ l
APS (10 %)	23 μ l	13 μ l
H ₂ O _{bidest}	2.222 ml	1.234 ml

Table 29: Resolving gel buffer (4-fold)

Components	Quantity
Tris (1.5 M)	45.5 g
SDS (0.4 %)	1 g
H ₂ O _{bidest}	<i>ad</i> 250 ml
HCl	to pH 8.8

Table 30: Stacking gel buffer (4-fold)

Components	Quantity
Tris (500 mM)	6.1 g
SDS (0.4 %)	0.4 g
H ₂ O _{bidest}	<i>ad</i> 100 ml
HCl	to pH 6.8

Purified protein samples were mixed 1 to 4 with SDS loading buffer (Table 32) and incubated at 95 °C for 5 min. In case of crude cell extracts, the heating time was extended to 10 minutes. To ensure a good comparability between the lines, the amount of protein in each sample was adjusted to a similar value. The gels were inserted into an electrophoresis chamber (Bio-Rad Laboratories, Munich, Germany) which was filled with 1 to 10 diluted electrophoresis buffer (Table 31). Of the prepared samples, up to 20 μ l (ideally about 3 μ g of protein) were carefully filled into the gel pockets, as well as a protein marker (PageRuler™ Unstained Protein Ladder, Thermo Fisher Scientific, Braunschweig, Germany) for size comparison. The electrophoresis was carried out at 100 V for the stacking gel and was increased up to 200 V, depending on the purity of the samples. Then, the gels were carefully removed from between the glass plates and placed in Coomassie stain (Table 33) for 30 minutes up to overnight, shaking gently. The destaining was carried out in water overnight or by using a destaining solution (Table 34).

Table 31: Electrophoresis buffer (10-fold)

Components	Quantity
Tris	250 mM
Glycine	1.92 M
SDS	1 %
H ₂ O _{bidest}	<i>ad</i> 1000 ml

Table 32: SDS loading buffer (4-fold)

Components	Quantity
Tris/HCl	200 mM
SDS	8 %
Glycerol	50 %
β-Mercaptoethanol	4 %
Bromphenole blue	0.04 %
H ₂ O _{bidest}	<i>ad</i> 50 ml

Table 33: Coomassie stain

Components	Quantity
Coomassie Brilliant Blue R250	0.2 %
Ethanol	40 %
Acetic acid	10 %
H ₂ O _{bidest}	<i>ad</i> 500 ml

Table 34: Destaining solution

Components	Quantity
Ethanol	40 %
Acetic acid	10 %
H ₂ O _{bidest}	<i>ad</i> 500 ml

2.7.6 Enzyme activity assay

The activity of the MCOs was measured by their ability to oxidize the following substrates: 2,2'-azinobis-3-ethylbenzothiazoline-6-sulfonic acid (ABTS), 4-hydroxy-3,5-dimethoxybenzaldehyde azine (syringaldazine) and 2-methoxyphenol (guaiacol).

The substrate stock solutions: ABTS 60 mM, syringaldazine 0.1 mM and guaiacol 10 mM, were prepared with water and stored at -20 °C, 4 °C and room temperature, respectively. For the reaction mixture, the stock solutions were diluted 1:20 with buffers of a pH that favored the activity of *T. versicolor* laccase: ABTS with pH 3.6 (acetate buffer), syringaldazine with pH 6.4 (phosphate buffer) and guaiacol with pH 4.0 (acetate buffer) (Eichlerová *et al.* 2012). The corresponding buffers were prepared by using the method and stock solutions described in chapter 2.2.

Then, the enzyme solution was blended 1:10 with the respective substrate reaction mixture and incubated for 20 min at room temperature. The absorption of the oxygenized product was measured at 420 nm (ABTS), 525 nm (syringaldazine) and 470 nm (guaiacol) with the corresponding reaction mixture as blank.

3 Results

3.1 *Ips typographus* sample collection and dissection

The *Ips typographus* bark beetle specimens were collected in a forest near Hannover (Germany) by using specific pheromone packages positioned inside slot traps. After capturing, the samples were immediately stored at -20 °C. Prior to dissection, the exterior of the beetles was disinfected with ethanol and they were fixated on paraffin plates with their ventral abdomen upwards (Figure 15).



Figure 15: *Ips typographus* fixated on paraffin plates prior to preparation. After disinfection with ethanol, the bark beetles were immobilized on petri dishes filled with paraffin with their ventral abdomen upwards.

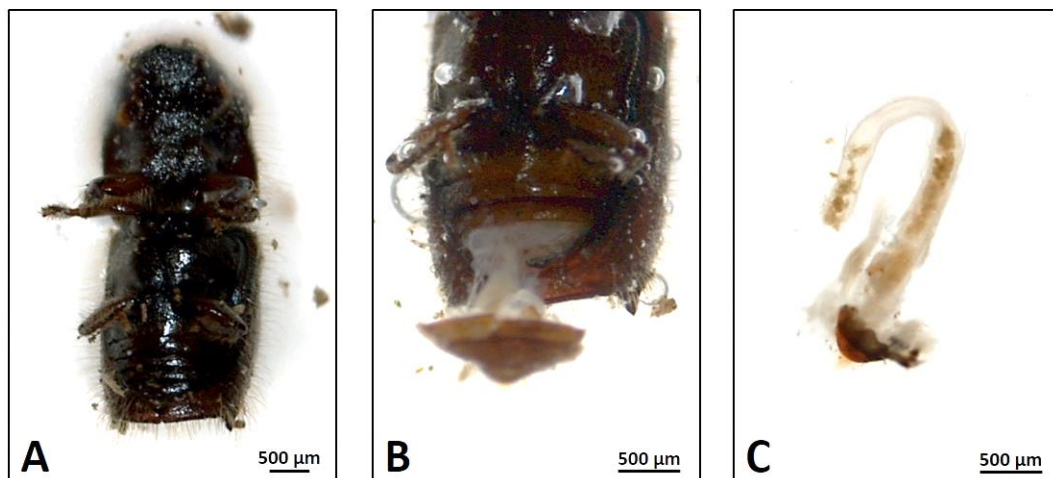


Figure 16: Dissection of the bark beetle intestines. The abdomen of the immobilized bark beetle (A) was opened with sterilized tweezers (B) and the intestines were removed (C), freed of shell pieces, and collected in sterile phosphate buffered saline prior to DNA extraction.

The immobilized beetles were submerged in sterile phosphate buffered saline. Their abdomen was opened with small tweezers, before the gastrointestinal tract was pulled out (Figure 16). After the removal of residual shell pieces, the intestines were collected in 1.5 ml sample tubes and stored at -20 °C.

3.2 DNA extraction and Illumina HiSeq Sequencing

The dissection of approximately 500 bark beetles yielded roughly 4.6 ng of metagenomic DNA with an average fragment length of 500 bp. After genomic library creation and Illumina HiSeq sequencing at the Heinrich Pette Institut (Hamburg, Germany), the raw data was trimmed and assembled (cf. chapter 2.6.1). The generated metagenomic dataset counted about 219 Mbp in total length, a number of 127909 predicted unique genes on 324359 contigs and a GC content of 41 %.

An initial taxonomic distribution of the whole metagenomic data (Figure 17) was processed by the MG-RAST metagenomics analysis server (generated 2014-12-09) (Meyer *et al.* 2008) with a Lowest Common Ancestor (LCA) algorithm and showed a ratio of Bacteria to Eukaryota of almost precisely 60 to 40 %.

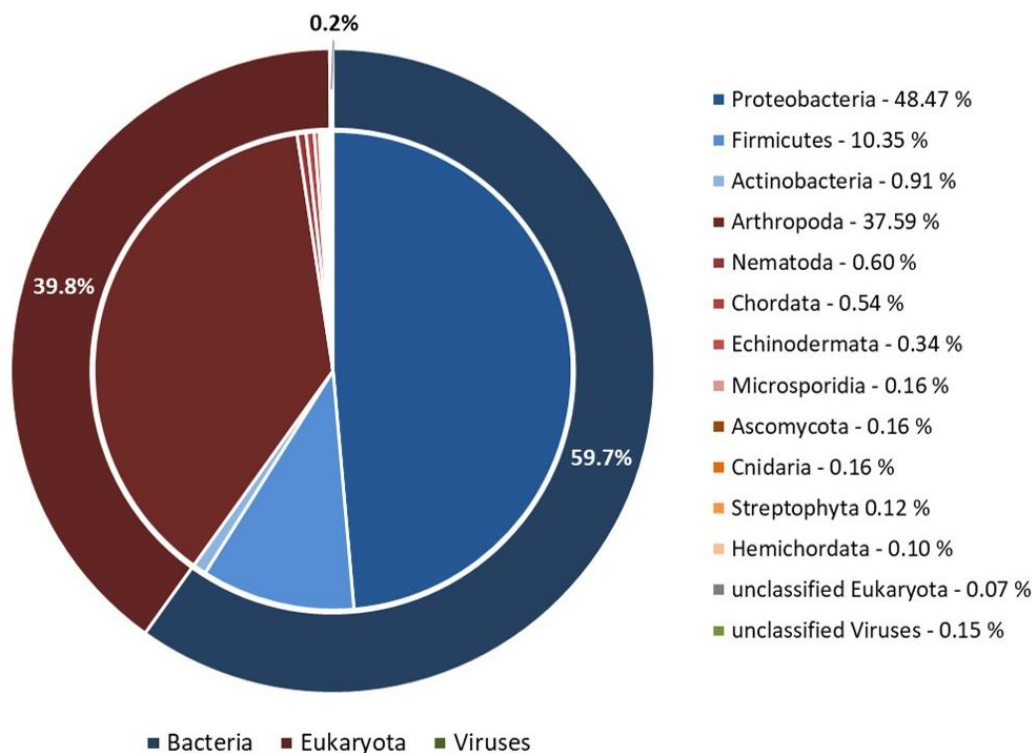


Figure 17: Taxonomic hits distribution of the whole raw *Ips typographus* metagenomic dataset. The complete Illumina HiSeq sequence data was uploaded to the MG-RAST metagenomics analysis server (Meyer *et al.* 2008) and computed with a Lowest Common Ancestor (LCA) algorithm to classify the phylum origin of every single sequence.

Nearly half of all sequences derived from Proteobacteria (48.5 %) and approximately 10 % were part of Firmicutes and 1 % of the Actinobacteria phylum. The largest part of eukaryotic sequences belonged to Arthropoda with 37.6 % while the other 2.2 % were associated with Nematoda (0.60 %), Chordata (0.54 %), Echinodermata (0.34 %), Microsporidia (0.16 %), Ascomycota (0.16 %), Cnidaria (0.16 %), Streptophyta (0.12 %) and Hemichordata (0.10 %). Only regarding the 59.7 % of bacterial sequences, Proteobacteria accounted for 81.2 %, Firmicutes for 17.3 % and Actinobacteria for 1.5 % of them.

3.3 Phylogenetic analysis by 16S rRNA gene sequencing

First, primer pairs for bacterial, archaeal, and eukaryotic ribosomal DNA sequencing were designed by combining specific binding sequences with Roche 454 pyrosequencing adaptors, key and identification codes. Several candidates were evaluated for performance and coverage of relevant taxonomic groups previously found in comparable wood-feeding insect gut samples *in silico*. This was carried out by computing them with the TestPrime online tool (version 1.0) (Klindworth *et al.* 2013) of the SILVA ribosomal RNA gene database project (SILVAngs version 1.2; www.arb-silva.de) (Quast *et al.* 2013). The most promising primer pairs (Table 8) were used to carry out the 16S and 18S rRNA gene PCR, with a template of metagenomic DNA isolated from the same beetle specimens, the *Ips typographus* dataset was created from (cf. chapter 2.6.2). Unfortunately, the archaeal primer pair did not yield any 16S rRNA gene amplicates.

After amplicon purification, the samples were sequenced by the Göttingen Genomics Laboratory and the resulting data was also processed by SILVAngs.

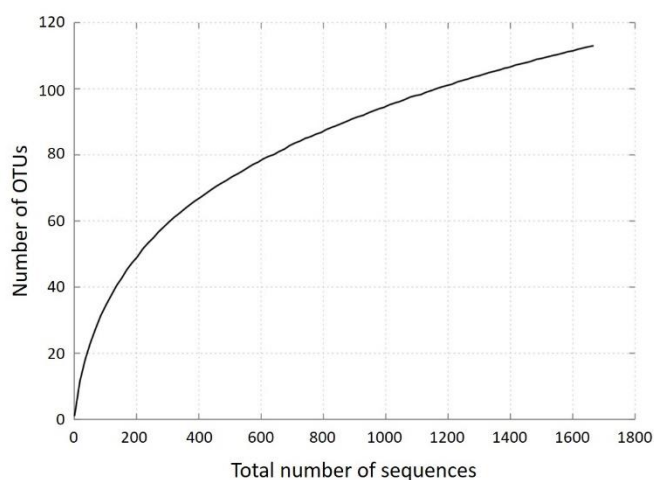


Figure 18: Rarefaction curve of 16S rRNA gene sequences from the *Ips typographus* gut microbiome. There were 1667 bacterial rRNA gene sequences found in total with 113 unique operational taxonomic units (OTUs). The rarefaction curve shows a distinctive flattening of the slope with a potential maximum approximating 140 OTUs.

There were 1667 bacterial rRNA gene sequences found with 113 unique operational taxonomic units (OTUs) distributed in 26 different genera. By displaying a distinct flattening of the slope, the rarefaction curve (Figure 18) indicated a potential maximum number of OTUs close to 140.

Table 35: Taxonomic distribution of *Ips typographus* metagenome 16S rRNA gene OTUs.

No. of OTUs	Percentage	Phylum	Class	Order	Genus
1	0.88 %	Actinobacteria	Actinobacteria	Propionibacteriales	<i>Propionibacterium</i>
2	1.77 %	Firmicutes	Bacilli	Bacillales	<i>Bacillus</i>
1	0.88 %	Firmicutes	Bacilli	Bacillales	<i>Geobacillus</i>
5	4.42 %	Firmicutes	Bacilli	Bacillales	<i>Planomicrobium</i>
1	0.88 %	Firmicutes	Bacilli	Lactobacillales	<i>Carnobacterium</i>
1	0.88 %	Proteobacteria	α -Proteobacteria	Rhizobiales	<i>Ochrobactrum</i>
1	0.88 %	Proteobacteria	α -Proteobacteria	Rhizobiales	<i>Phyllobacterium</i>
1	0.88 %	Proteobacteria	α -Proteobacteria	Rhodobacterales	<i>Paracoccus</i>
4	3.54 %	Proteobacteria	γ -Proteobacteria	B38	-
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Cedecea</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Cronobacter</i>
27	23.89 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Enterobacter</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Erwinia</i>
9	7.96 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Escherichia-Shigella</i>
7	6.19 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Hafnia</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Leminorella</i>
36	31.86 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Morganella</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Pantoea</i>
3	2.65 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Providencia</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Rahnella</i>
3	2.65 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Raoultella</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Serratia</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Enterobacteriales	<i>Yersinia</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Pseudomonadales	<i>Acinetobacter</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Pseudomonadales	<i>Enhydrobacter</i>
1	0.88 %	Proteobacteria	γ -Proteobacteria	Pseudomonadales	<i>Pseudomonas</i>

The taxonomic distribution of the 16S rRNA gene OTUs (Table 35) showed that Actinobacteria with one single OTU accounted for 0.88 % of the sequences while 9 (7.96 %) belonged to Firmicutes and 103 (91.15 %) to Proteobacteria. Two of the OTUs contributed by Firmicutes were from *Bacillus* and five from *Planomicrobium* while *Carnobacterium* and *Geobacillus* each supplied one sequence. *Morganella* and *Enterobacter* were by far the best represented Proteobacteria with 36 and 27 OTUs corresponding to 31.86 % and 23.89 %, respectively. The genera *Escherichia-Shigella* provided 9 OTUs, *Hafnia* 7, *Providencia* 3 and *Raoultella* also 3. In addition, four sequences also originated from the uncharacterized Proteobacterium B38 while the 14 remaining proteobacterial genera only accounted for one single OTU. Overall, this summed the number of non-abundant taxa of < 1 % diversity up to 17, representing 15.04 % of all OTUs.

Regarding the whole 1667 16S rRNA gene sequences classified by SILVAngs, on phylum level (Figure 19A) there were 1625 sequences from Proteobacteria (97.48 %), 41 from Firmicutes (2.46 %) and a single one from Actinobacteria (0.06 %) detected.

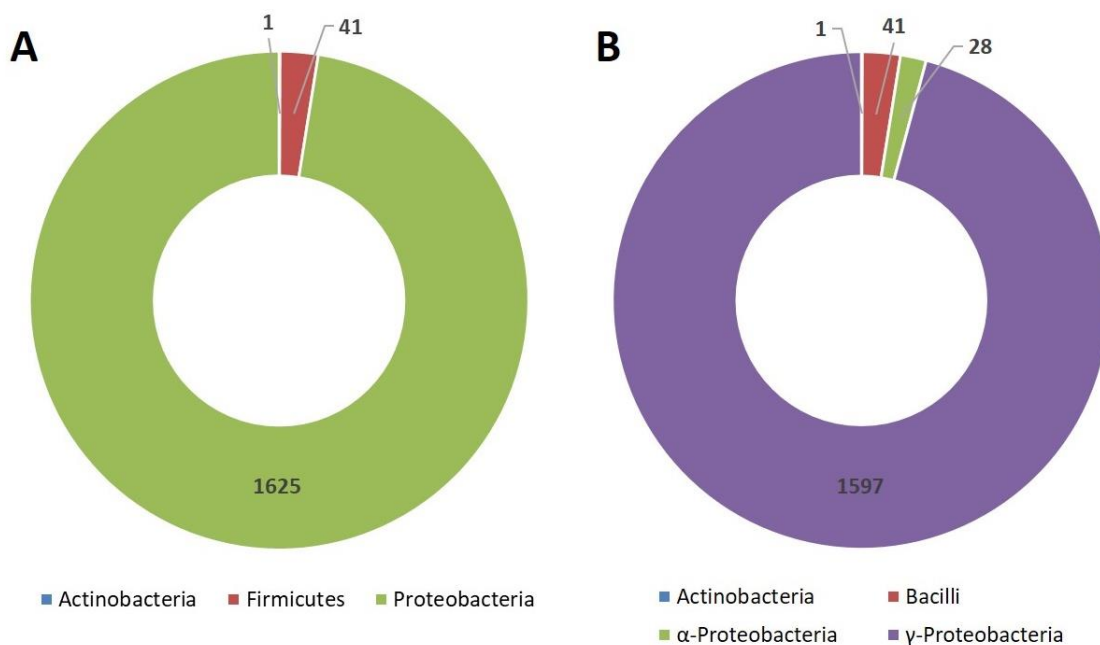


Figure 19: Taxonomic classification of the *Ips typographus* bacterial 16S rRNA gene sequences on phylum and class level. After 16S rRNA gene PCR and purification, the amplicons were sequenced by the Göttingen Genomics Laboratory. The resulting 1667 sequences were processed by SILVAngs (version 1.2) (Quast *et al.* 2013). On phylum level (A), 1625 sequences belonging to Proteobacteria, 41 to Firmicutes and 1 to Actinobacteria were discovered. On class level (B), the 41 Firmicutes sequences were all assigned to Bacilli while the ratio of Alpha- to Gammaproteobacteria was 1 to 57 (28 to 1597).

A deeper differentiation (Figure 19B) showed that the ratio of α - to γ -Proteobacteria was 1 to 57.

Overall, the taxonomic classification on genus level (Figure 20) revealed only 26 different genera adding to the 1667 sequences with 21 being Proteobacteria, 4 Firmicutes and one Actinobacteria. *Morganella* supplied 57.11 % of all sequences while *Enterobacter* ranked second with 22.20 % and *Serratia* following with 5.64 %. Eight genera only contributed one single sequence while most of them accounted for less than 1 % in total.

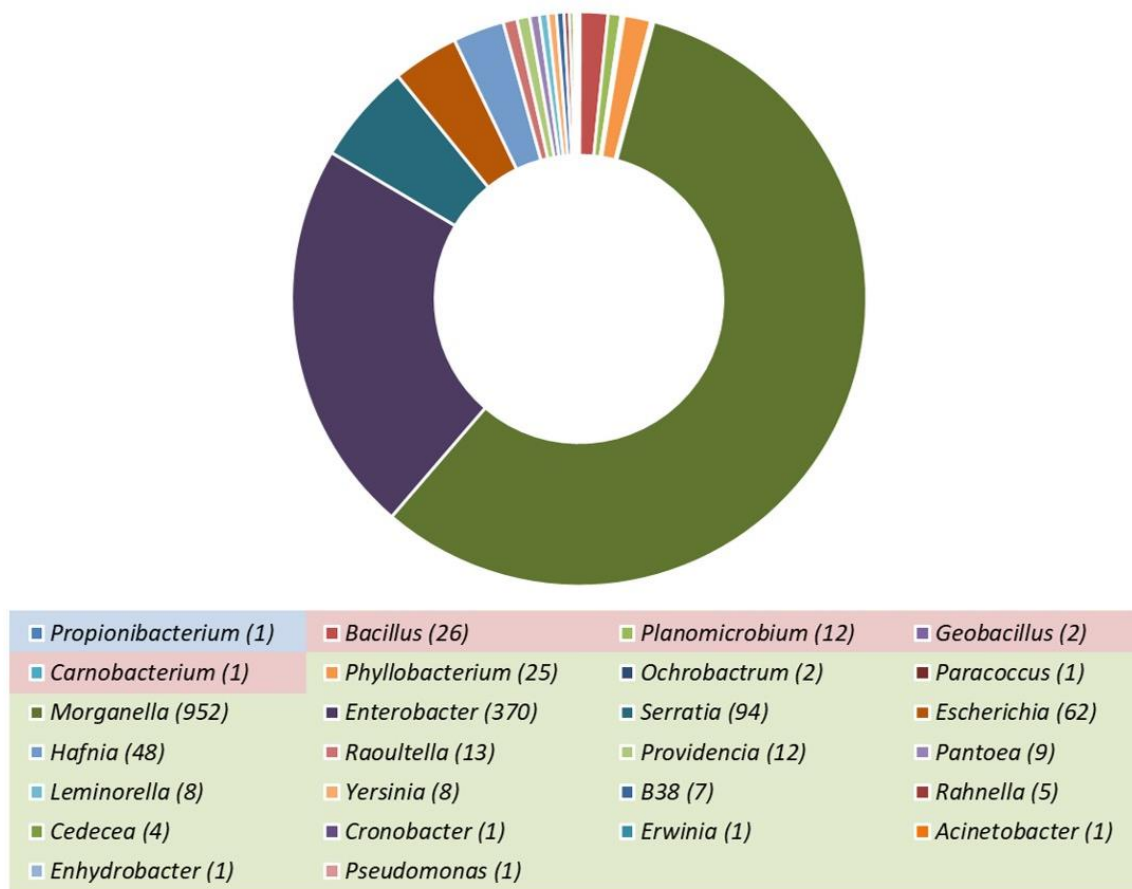


Figure 20: Taxonomic classification of the *Ips typographus* bacterial 16S rRNA gene sequences on genus level. After 16S rRNA gene PCR and purification, the amplicons were sequenced by the Göttingen Genomics Laboratory. The resulting 1667 sequences were processed by SILVAngs (version 1.2) (Quast *et al.* 2013). Phylum association of the genera were visualized by highlight color with Actinobacteria in blue, Firmicutes in red and Proteobacteria in green. With a number of 952 sequences, 57.11 % were associated with *Morganella* while *Enterobacter* came second with 370. Several other Proteobacteria contributed between 4 and 94 sequences whereas *Cronobacter*, *Erwinia*, *Acinetobacter*, *Enhydrobacter* and *Pseudomonas* each only accounted for one single sequence.

3.4 Phylogenetic analysis by 18S rRNA gene sequencing

The rarefaction curve of the 18S rRNA gene sequences of the *Ips typographus* gut microbiome (Figure 21) shows a practically flat slope indicating that nearly all possible OTUs were covered. There were 4250 classifiable 18S rRNA gene sequences found in total which resulted in 32 unique OTUs. Minus *Ips typographus* and plant derived sequences, there still were 6 potentially relevant OTUs left. Results were summarized in Table 36. Two OTUs were derived from the nematode *Micoletzkyia* while one originated from the oomycete genus *Phytophthora*. Three sequences belonged to fungi, one Basidiomycota associated with human skin surface and two Ascomycota, identified as *Wickerhamomyces* and Sarcosomataceae.

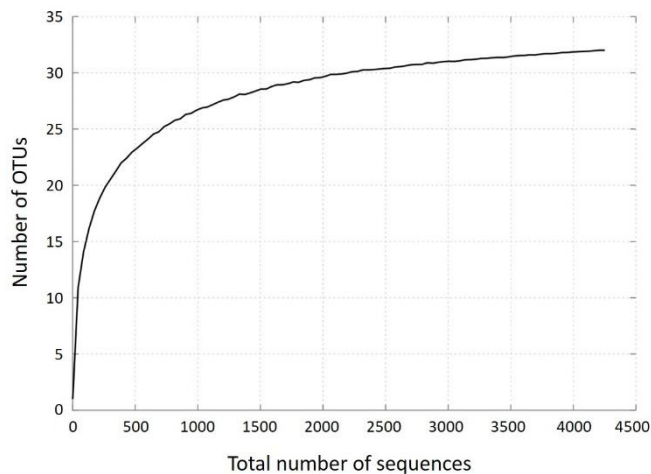


Figure 21: Rarefaction curve of 18S rRNA gene sequences from the *Ips typographus* gut microbiome. There were 4268 eukaryotic rRNA gene sequences found in total with 32 unique operational taxonomic units (OTUs). The rarefaction curve shows a nearly flat slope with a potential maximum number of OTUs of less than 35.

Table 36: Significant *Ips typographus* metagenome 18S rRNA gene OTUs minus insect and plant sequences.

No. of OTUs	Phylum/Division	Family	Genus	Potential function/origin
2	Nematoda	Diplogastridae	<i>Micoletzkyia</i>	Potential symbiont
1	Ascomycota	Phaffomycetaceae	<i>Wickerhamomyces</i>	Potential symbiont
1	Ascomycota	Sarcosomataceae	-	Rotten wood
1	Basidiomycota	-	-	Human skin surface
1	Heterokontophyta	Peronosporaceae	<i>Phytophthora</i>	Plant pathogen/ potential symbiont

3.5 Annotation by the DOE-JGI Microbial Annotation Pipeline

The *Ips typographus* metagenomic dataset provided by the Heinrich Pette Institute (Hamburg, Germany) was uploaded to the Genomes Online Database (GOLD; (Kyrpides 1999) annotated by the DOE-JGI Microbial Annotation Pipeline (Hunte-mann *et al.* 2015) and stored in the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012) for further analysis. This pipeline computed the annotation with the following steps: It started with the detection of RNA genes with the help of several RNA databases. Protein-coding genes were identified by comparison with known genes from the COG and Pfam databases and associated with Kyoto Ency-clopedia of Genes and Genomes (KEGG) orthology terms (Kanehisa *et al.* 2002). Afterwards, the genome was integrated into IMG by computing protein sequence similarities between its genes and genes of all other genomes in the system (Mar-kowitz *et al.* 2014).

After annotation, the *Ips typographus* dataset still comprised of 218.6 Mbp with 432305 protein-coding and 1915 RNA genes. The IMG database provided a phylo-genetic distribution based on best BLAST hits of protein-coding genes (Figure 22).

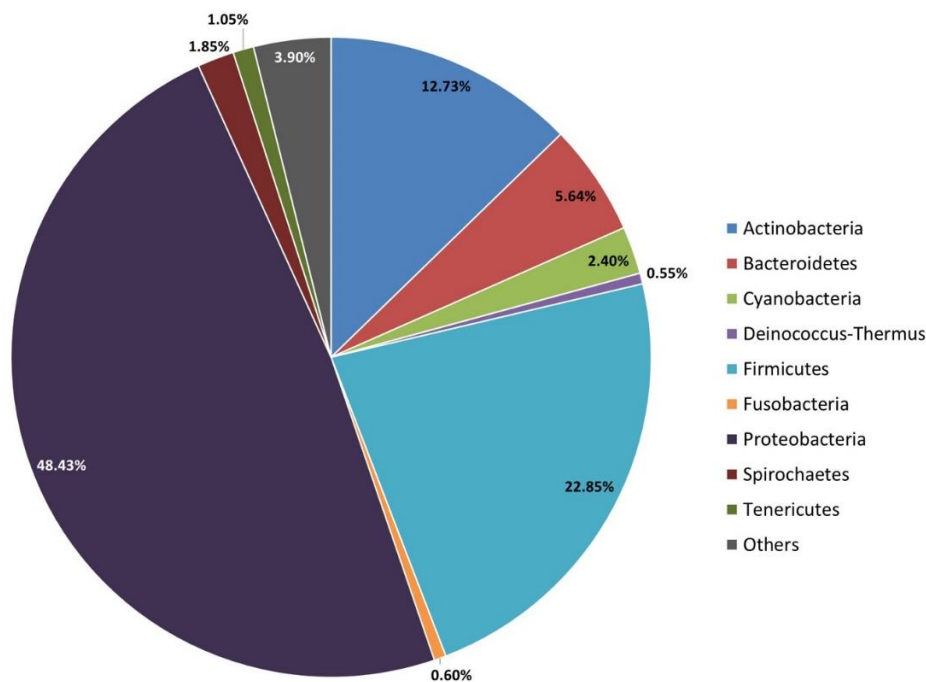


Figure 22: Phylogenetic distribution of bacterial genes in the *Ips typographus* dataset (IMG ID 56273). Annotation by the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012). Genome count by best Blast hits at 30 % identity with only phyla above 0.5 % shown.

At 30 % BLAST identity, the *Ips typographus* dataset was divided into the 9 most represented phyla while the ones below 0.5 % were summarized as “others”. Almost half of all protein-coding genes were connected to Proteobacteria, nearly a fourth to Firmicutes and an eighth to Actinobacteria. Bacteroidetes provided 5.6 % of the sequences and Cyanobacteria, Spirochaetes, Tenericutes, Fusobacteria and Deinococcus-Thermus followed with 2.4 %, 1.85 %, 1.05 %, 0.60 % and 0.55 %, respectively. Protein-coding sequences associated with other phyla still accounted for 3.90 %.

The largest contig counted 444973 bp and a search with the Basic Local Alignment Search Tool (BLAST; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Camacho *et al.* 2009) showed an association with the *Erwinia* genus. Of the 50 longest contigs and consequently 7.7 Mbp of the whole dataset, 44 sequences were linked to Enterobacteriaceae while 6 were of Firmicutes or more specifically of *Bacillus* origin.

3.6 Comparison with other gut-associated metagenomic datasets

To relate the phylogenetic gene distribution of *Ips typographus* to other insect and vertebrate metagenomes, corresponding gastrointestinal datasets were chosen randomly from the IMG database (Figure 23). Two of the datasets derived from wood-feeding beetles as well, the first one from the Asian long-horned beetle *Anoplophora glabripennis* (IMG ID 1835) with a length of 281 Mbp and the second from the Passalidae beetle *Viturius sinuatocollis* (IMG ID 12352) with 143 Mbp. Three other datasets also originated from insects: from the termite *Nasutitermes corniger* (IMG ID 59112) with 210 Mbp, the woodwasp *Sirex noctilio* (IMG ID 2240) with 178 Mbp and the European honeybee *Apis mellifera* (IMG ID 9040) with 250 Mbp. The last three metagenomes originated from vertebrates. Amongst them were the domestic chicken *Gallus gallus* (IMG ID 45592) with 286 Mbp length, the dog *Canis familiaris* (IMG ID 680) with 241 Mbp and the giant panda *Ailuropoda melanoleuca* (IMG ID 65432) with 18 Mbp.

Interestingly, *Ips typographus* showed the second largest number of genes assigned to Proteobacteria (48.4 %) and the second lowest count of Firmicutes (22.9 %) and Fusobacteria (0.6 %), each time in direct pursuit of *Sirex noctilio* with 61.3 %, 18.0 % and 0.4 %, respectively.

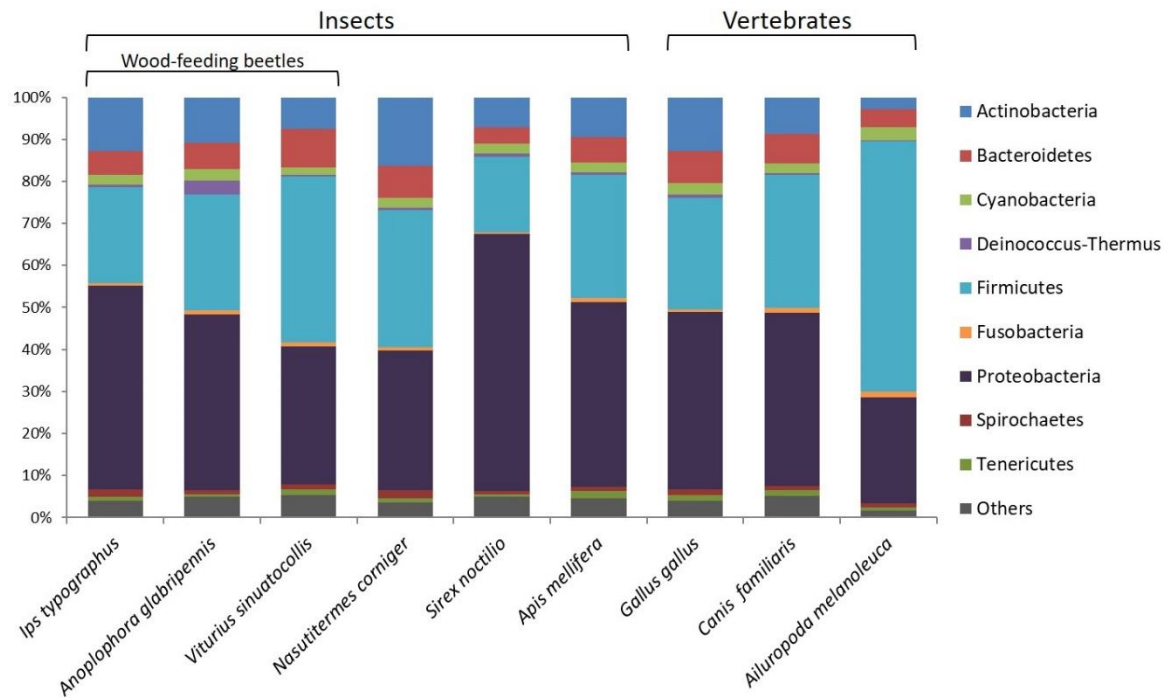


Figure 23: Comparison of phylogenetic distribution between *Ips typographus* and eight other gut-associated metagenomic datasets. The phylogenetic distribution of protein-coding genes was compared between *Ips typographus* (IMG ID 56273) and eight other gut-associated datasets. Two of them were from wood-feeding beetles as well, one from the Asian long-horned beetle *Anoplophora glabripennis* (IMG ID 1835) and one from the Passalidae beetle *Viturius sinuaticollis* (IMG ID 12352). Three other datasets were also derived from insects: from the termite *Nasutitermes corniger* (IMG ID 59112), the woodwasp *Sirex noctilio* (IMG ID 2240) and the European honeybee *Apis mellifera* (IMG ID 9040). The last three metagenomes originated from vertebrates. Amongst them were the domestic chicken *Gallus gallus* (IMG ID 45592), the dog *Canis familiaris* (IMG ID 680) and the giant panda *Ailuropoda melanoleuca* (IMG ID 65432). Annotation by the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012). Genome count by best Blast hits at 30 % identity with only phyla above 1 % for one of the datasets shown. Percentages shown in Table 44 in the appendix section.

The bark beetle also had the second highest rate of genes associated with Spirochaetes at 1.85 % and Actinobacteria at 12.7 %, the latter along with *Gallus gallus*, only outnumbered by *Nasutitermes corniger* with 1.92 % and 16.4 %, each. Regarding Bacteroidetes, Cyanobacteria, Deinococcus-Thermus and Tenericutes, *Ips typographus* scored average numbers in comparison to the other datasets with 5.6 %, 2.4 %, 0.6 % and 1.1 %, respectively.

To assess *Ips typographus* and the same previously reviewed datasets on a functional basis, the protein-coding genes annotated by IMG and distributed into highest ranking KEGG groups (Kanehisa *et al.* 2002) were visualized in Figure 24. The bark beetle possessed comparably numerous genes associated with metabolic functions. It ranked highest overall in the categories for AA, carbohydrate, energy, lipid as well as cofactor and vitamin metabolism with 15.7 %, 17.2 %, 9.1 %, 5.1 % and 8.9 %, respectively.

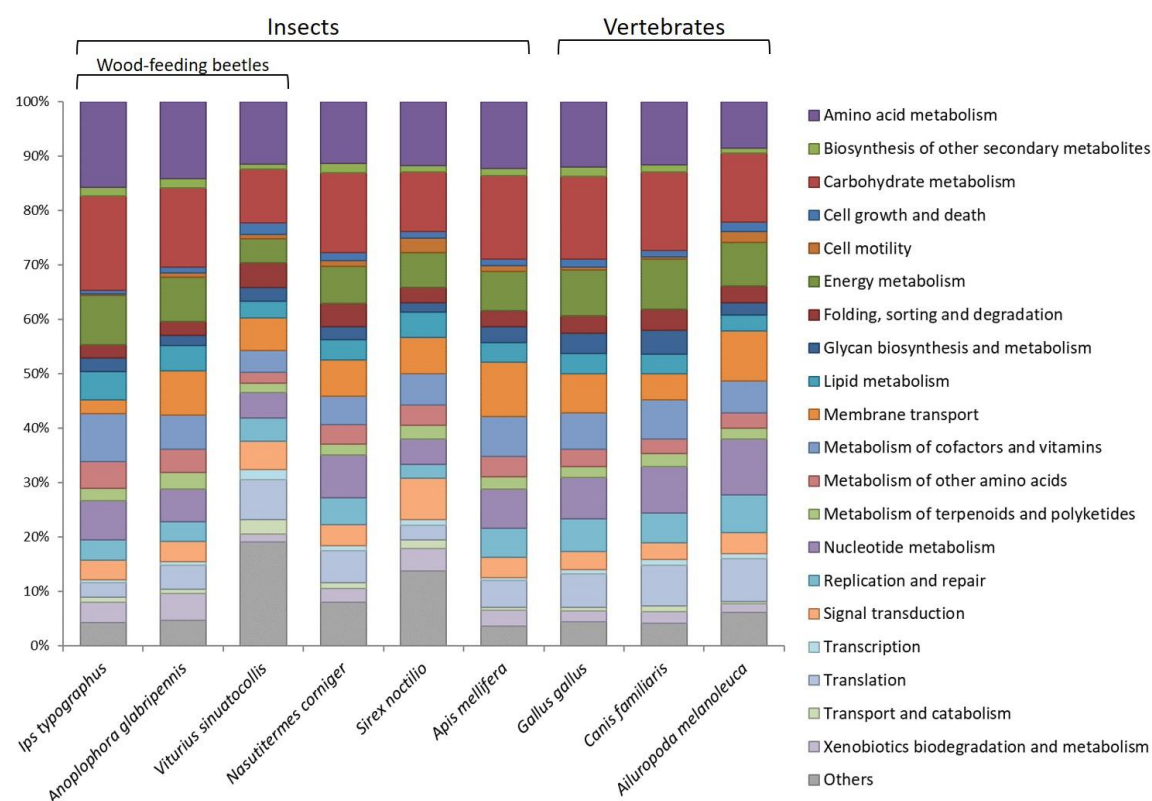


Figure 24: Comparison of gene functional distribution according to KEGG between *Ips typographus* and eight other gut-associated metagenomic datasets. The functional distribution of protein-coding genes according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.* 2002) (Kanehisa *et al.* 2002) was compared between *Ips typographus* (IMG ID 56273) and eight other gut-associated datasets. Two of them were from wood-feeding beetles as well, one from the Asian long-horned beetle *Anoplophora glabripennis* (IMG ID 1835) and one from the Passalidae beetle *Viturinus sinuaticollis* (IMG ID 12352). Three other datasets were also derived from insects: from the arboreal termite *Nasutitermes corniger* (IMG ID 59112), the woodwasp *Sirex noctilio* (IMG ID 2240) and the European honeybee *Apis mellifera* (IMG ID 9040). The last three metagenomes originated from vertebrates. Amongst them were the domestic chicken *Gallus gallus* (IMG ID 45592), the dog *Canis familiaris* (IMG ID 680) and the giant panda *Ailuropoda melanoleuca* (IMG ID 65432). Annotation by the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012). Percentages shown in Table 45 in the appendix section.

3.7 Novel putative multicopper oxidases

3.7.1 Enzymes originating from the *Ips typographus* dataset

For the identification of novel MCOs in the *Ips typographus* dataset, a search with “multicopper oxidase” as keyword was processed in the IMG metagenome. Resulting gene hits were checked for their completeness and the presence of all four CBS motifs. Six candidate genes met these requirements and were displayed in Table 37 while their nucleotide and AA sequences were listed in the corresponding appendix section.

Table 37: Putative MCO genes found by *in silico* screening of the *Ips typographus* metagenome

Putative MCO gene	IMG Gene ID	Gene Product name	Gene size [nt]
<i>ips1282</i>	Ga0063521_1000001282	putative multicopper oxidase	1665
<i>ips2204</i>	Ga0063521_10002204	putative multicopper oxidase	1629
<i>ips14138</i>	Ga0063521_100014138	putative multicopper oxidase	1575
<i>ips21622</i>	Ga0063521_100021622	putative multicopper oxidase	1608
<i>ips24328</i>	Ga0063521_100024328	putative multicopper oxidase	1605
<i>ips28714</i>	Ga0063521_100028714	putative multicopper oxidase	1611

Table 38: BLAST results and potential origin of the putative *Ips typographus* MCO genes

Putative MCO gene	Genus	Family	Query cover	Identity
<i>ips1282</i>	<i>Providencia</i>	Enterobacteriaceae	100 %	81 %
<i>ips2204</i>	<i>Erwinia</i>	Enterobacteriaceae	56 %	74 %
<i>ips14138</i>	<i>Cedecea</i>	Enterobacteriaceae	100 %	79 %
<i>ips21622</i>	<i>Morganella</i>	Enterobacteriaceae	100 %	92 %
<i>ips24328</i>	<i>Rahnella</i>	Enterobacteriaceae	99 %	89 %
<i>ips28714</i>	<i>Rouxiella</i>	Enterobacteriaceae	99 %	92 %

<https://blast.ncbi.nlm.nih.gov/Blast.cgi> access 2018-07-29 (Altschul *et al.* 1990)

Putative metagenomic MCO genes were computed with the National Center for Biotechnology Information's (NCBI's) Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990), to determine the organisms with the most similar nucleotide sequences and therefore the potential origin of the genes (Table 38).

All six discovered metagenomic MCO genes possessed the highest identity to known nucleotide sequences of the family Enterobacteriaceae. *Ips1282*, *ips14138* and *ips21622* all covered 100 % of the sequence length of the genes they were most similar to. They shared the highest identity with genes of the genera *Providencia* (81 %), *Cedecea* (79 %) and *Morganella* (92 %), respectively. *Ips24328* featured the highest identity to a gene from *Rahnella* (89 %) and *ips28714* was most similar to a gene from *Rouxiella* (92 %) while both presented a query cover of 99 %. The putative MCO that resembled known genes from the NCBI database the least was

ips2204. It only covered 56 % of the sequence of an identified gene from *Erwinia* and showed 74 % similarity to it.

Comparing the DNA sequences of these six metagenomic genes in relation to *E. coli*'s *cueo* resulted in the phylogenetic tree shown in Figure 25.

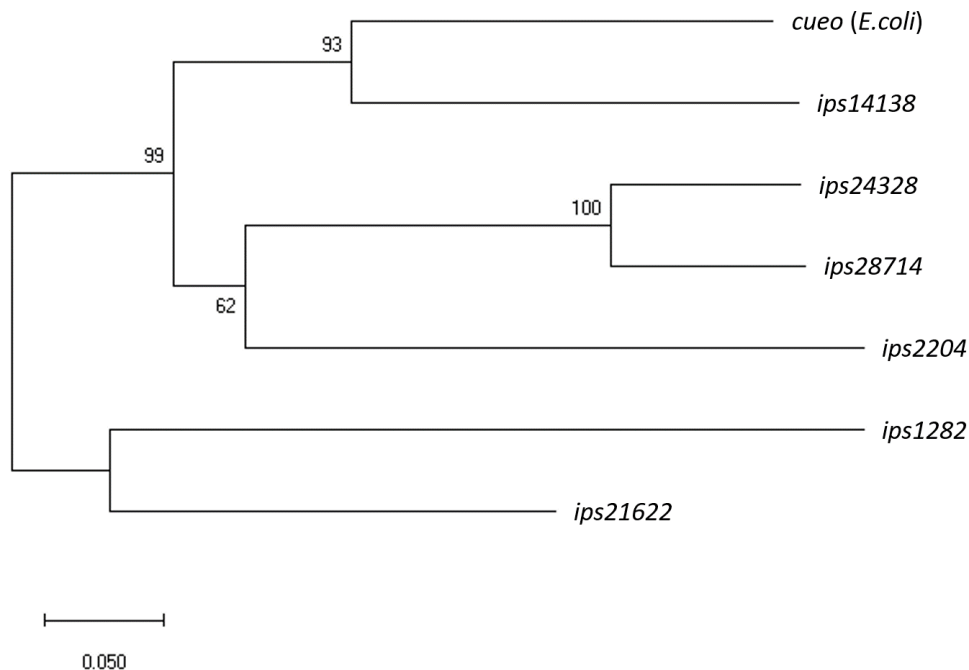


Figure 25: Phylogenetic tree of the putative *Ips typographus* MCO genes and *cueo* from *E. coli*. The similarity of the sequences correlates with the branch lengths that is measured in the number of substitutions per site. Next to the branches, the percentage of trees in which the associated taxa clustered together is shown. The construction was carried out by using the Multiple sequence alignment tool M-Coffee (Notredame *et al.* 2000; Wallace *et al.* 2006) for sequence alignment and MEGA X (Kumar *et al.* 2018) for the phylogenetic analysis (statistical method: maximum likelihood; bootstrap replications: 100).

The alignment was done with the online tool T-Coffee (Notredame *et al.* 2000) which uses multiple alignment methods at the same time (M-Coffee: e.g. Clustal, Mafft, Probcons, Muscle). The phylogenetic relation was analyzed and visualized by using the software MEGA X (Kumar *et al.* 2018) with the maximum likelihood method and 100 bootstrap replications. The tree showed that *ips24328* and *ips28714* were the closest associated genes in all the calculations. In 62 % of the computed replications, *ips2204* clustered together with these two genes although the branches were quite long. *Ips14138* showed the highest similarity to *cueo* and both groups were clustered together in 99 % of the calculations. *Ips1282* and *ips21622* formed a cluster in 99 % of the computed trees, which was located the most distant from the other five gene sequences.

CueO	69	EETTH	HWH	GLEVP	81	109	ATCWF	HPH	QHGKT	121
lps1282	98	EDTTV	HWH	GLEIS	110	138	ATCWF	HPH	THGKT	150
lps2204	97	EATTI	HWH	GLALP	109	137	ATCWF	HPH	QHGRT	149
lps14138	96	EETTV	HWH	GLEVP	108	136	ATCWF	HPH	QHGKT	148
lps21622	98	EATTV	HWH	GLEIS	110	138	ATCWF	HPH	THGKT	150
lps24328	97	VASTV	HWH	GLEIP	109	137	STCWF	HPH	PHQTS	149
lps28714	97	VASTV	HWH	GLEIP	109	137	ATCWF	HPH	PHQTS	149
CueO	411	GDMML	HPFH	HGTQFR	426	469	YMA	HCH	LLEHEDTG	MML 485
lps1282	475	GDMML	HPFHV	HGTRFR	490	533	YMA	HCH	LLEHEDTG	MML 549
lps2204	462	GDGML	HPFH	HGAQFR	477	520	YMA	HCH	LLEHEDTG	MML 536
lps14138	445	GDMML	HPFH	HGTQFR	460	503	YMA	HCH	LLEHEDTG	MML 519
lps21622	455	GDMML	HPFH	HGTRFR	470	513	FMA	HCH	LLEHEDTG	MMT 529
lps24328	454	GDMML	HPFH	HGTQFR	469	512	YMA	HCH	LLEHEDTG	MML 528
lps28714	456	GDMML	HPFH	HGTQFR	471	514	YMA	HCH	LLEHEDTG	MML 530

Figure 26: Comparison of the four copper binding motifs of the putative MCOs from the *Ips typographus* metagenome and CueO from *E. coli*. The four copper binding motifs (framed in green) were aligned and illustrated with their respective AA positions. The AAs between the histidines can vary except for the first cysteine of the type 1 motif H-C-H-x(3)-H-x(4)-[MFL] (Reiss *et al.* 2013).

The comparison of the *Ips typographus* proteins' copper binding motifs (CBM) (Figure 26) showed that there were not many differences in the sequences surrounding the CBMs. The histidine motifs of CueO were always located upstream of their matching counterparts in the bark beetle proteins. The only direct deviation in CBM sequences was found in the third copper binding motif of lps1282, where the isoleucine between second and third histidine was substituted by valine. Overall, the CBMs of lps1282 were also consistently positioned the furthest downstream compared to the other proteins. Regarding AAs surrounding the CBM, the similarities between the sequences of lps24328 and lps28714 were particularly noticeable. Apart from single AA deviations, the most diverse sequence was observed in lps2204. Both in CBM one and three, there were unique AAs only present in lps2204 at this position. Directly upstream of the first histidine of CBM one, there was an isoleucine located instead of a leucine for CueO or a valine for the other four MCO. Also surrounding CBM one, the third and fourth AA downstream of the motif were alanine and leucine in place of primarily glutamic acid and isoleucine or valine. For CBM three, the distinct AAs of lps2204 were located in positions 464 and 474 with glycine present instead of methionine and alanine substituted for threonine in contrast to all the other sequences.

3.7.2 Enzymes originating from other bacterial strains

To identify specific qualities of the enzymes originating from the bark beetle intestines and to act as comparison, several available bacterial genomes were also screened for novel MCO genes. The MOTIF search tool of the GenomeNet network (Kanehisa 1997) and the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012) were used for this purpose. *Duganella* sp. HH01 was chosen because related strains were isolated from several aquatic and environmental habitats such as sewage, agricultural and forest soil where lignin is omnipresent (Hornung *et al.* 2013; Haack *et al.* 2016). *Stenotrophomonas maltophilia* K279a (Avison *et al.* 2000) belongs to the order Xanthomonadales and is closely related to the plant pathogens of *Xanthomonas* (Crossman *et al.* 2008). The plant symbiont *Sinorhizobium fredii* NGR234 was considered because of its ability to inhabit a broad range of different host plants (Trinick 1980; Schmeisser *et al.* 2009).

The screening was carried out by searching for the copper binding site (CBS) type 1 motif H-C-H-x(3)-H-x(4)-[MFL] (Reiss *et al.* 2013) in the proteome of organisms that were represented in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa *et al.* 2002) and shared the same genus as the target species. Protein sequences that were found to contain the CBM were each compared via BLAST to the above-mentioned target genomes in the IMG database.

Table 39: Putative MCO genes found by *in silico* screening of several bacterial genomes: *Duganella* sp. HH01, *S. maltophilia* K279a and *S. fredii* NGR234 in the Integrated Microbial Genomes database (IMG)

Source organism	Putative MCO gene	IMG Gene ID	Annotated gene product name	Gene size [bp]
<i>Duganella</i> sp. HH01	<i>hh01-218</i>	2523500218	multicopper oxidase	1194
<i>Duganella</i> sp. HH01	<i>hh01-482</i>	2523500482	putative multicopper oxidase	1392
<i>S. maltophilia</i> K279a	<i>sm3532</i>	642693532	putative copper resistance protein	1869
<i>S. maltophilia</i> K279a	<i>sm4696</i>	642694696	putative copper resistance protein/multicopper oxidase	1809
<i>S. fredii</i> NGR234	<i>ngr147</i>	643825147	putative multicopper oxidase	1401
<i>S. fredii</i> NGR234	<i>ngr578</i>	643822578	putative copper-containing oxidase	1350
<i>S. fredii</i> NGR234	<i>ngr600</i>	643822600	potential multicopper oxidase	1491
<i>S. fredii</i> NGR234	<i>ngr688</i>	643822688	multicopper oxidase	1350

The protein sequences with the highest identities were screened for the other 3 CBMs and if found, further analyzed. Putative viable MCOs that were identified with this method were listed in Table 39.

Two candidate genes – *hh01-218* (1194 bp) and *hh01-482* (1392 bp) – were found in the genome of *Duganella* sp. HH01. These were also both already annotated as (putative) MCOs by IMG. The genes *sm3532* (1869 bp) and *sm4696* (1809 bp) were derived from *S. maltophilia* K279a and fulfilled the requirements as well. Both were classified as putative copper resistance proteins whereas *sm4696* was also marked as MCO.

The screening of *S. fredii* NGR234 produced four complete putative MCO genes: *ngr147* (1401 bp), *ngr578* (1350 bp), *ngr600* (1491 bp) and *ngr688* (1350 bp). With exception of *ngr578*, which was annotated as copper-containing oxidase, the genes were classified as (putative/potential) MCOs by IMG. It was noticeable that the nucleotide sequences of *ngr578* and *ngr688* had a query cover of 100 % and were 91 % identical to each other. Their protein sequences showed an even larger similarity with 95 %.

Comparing the copper binding motifs of the genomic MCOs (Figure 27), it was noticeable that the AA sequences of NGR578 and NGR688 were completely identical in the area surrounding the CBMs. They also showed more resemblance to the HH01 proteins than to NGR147 and NGR600. All eight proteins contained four copper binding sites (CBS) except for NGR147, which had an additional motif between AA positions 62 and 74. As this motif possessed a threonine in the center, the equivalent to the ones from the other proteins seemed to be located between AA 92 and 104 with histidines flanking tryptophan.

The neighborhood of these putative MCO genes (Figure 28) consisted of ORFs with various predicted functions. Like the genetic sequences themselves, the surrounding regions of *ngr578* and *ngr688* showed a high similarity to each other. Except for *ngr147*, all MCOs had genes coding for proteins with copper binding and/or transport function in the immediate vicinity. Neighboring genes of *hh01-482* and *ngr147* were also linked to carbohydrate binding and transport. Whereas *hh01-218* and *hh01-482* as well as *ngr578* and *ngr688* were located close to ORFs associated with membrane binding and transport across.

HH01-218	57	EHTSV HW HGQRLP	69	99	GTFMY HP HADEMA	111
HH01-482	123	EHTSV HW HGQRLP	135	165	GTFMY HP HADEMT	177
SM3532	112	ADTSI HW HGIILP	124	154	GTYWY HS HSGFQE	166
SM4696	113	HPTSI HW HGILLP	125	155	GTYWY HS HSMFQE	167
NGR147	62	LGDAV HT HLPPVL	74	134	GTFWY HP HCNTLT	146
	92	EPTTV HW HGLRIA	104			
NGR578	122	EHTTI HW HGMILP	134	164	GTFMY HP HSDEM V	176
NGR600	86	EETLI HW HGLTPP	98	128	GTNWM HS HHLQEQ	140
NGR688	122	EHTTI HW HGMILP	134	164	GTFMY HP HSDEM V	176
HH01-218	192	LTMTN HP IHLHGHEFF	207	247	WAF HCH KSHHTMNAMGH	263
HH01-482	258	LTMTN HP IHHGHEFF	273	313	WAF HCH KSHHTMNAMGH	329
SM3532	550	DTMMT HP IHLHGMWSD	565	600	WAF HCH LLYHMEAGMMR	616
SM4696	530	DTMMQ HP IHLHGVWSD	545	580	WAY HCH LLYHMEAGMMR	596
NGR147	393	RTPHA HP IHLHGLSFR	408	445	WVI HCH IEHQKTGMTG	461
NGR578	257	LTMTN HP IHMHG YDFE	272	312	WAI HCH KSHHTMNAMGH	328
NGR600	422	HTTMA HP MHLHGHHFQ	437	474	WAI HCH HLYHMNGGMMT	490
NGR688	257	LTMTN HP IHMHG YDFE	272	312	WAI HCH KSHHTMNAMGH	328

Figure 27: Comparison of the four copper binding motifs of the putative MCOs from *Duganella* sp. HH01, *S. maltophilia* K279a and *S. fredii* NGR234. The four copper binding motifs (framed in green) were aligned and illustrated with their respective AA positions. The AAs between the histidines can vary except for the first cysteine of the type 1 motif H-C-H-x(3)-H-x(4)-[MFL] (Reiss *et al.* 2013).

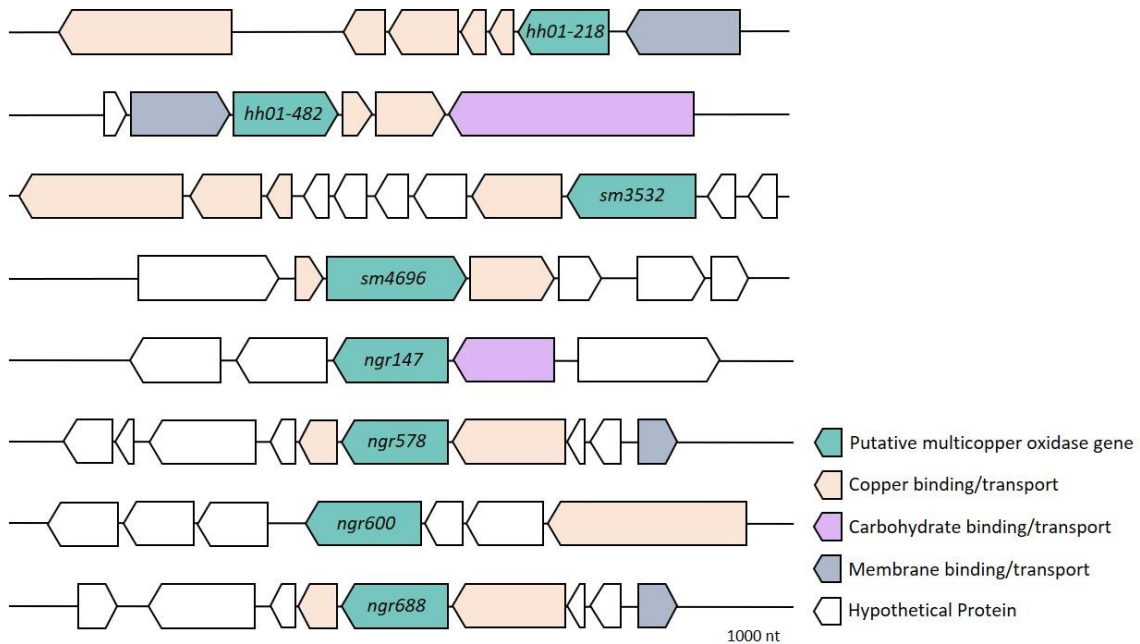


Figure 28: ORF neighborhood of the putative MCO Genes found by *in silico* screening of *Duganella* sp. HH01, *S. maltophilia* K279a and *S. fredii* NGR234. The putative MCO genes (green) are surrounded by ORFs with predicted functions linked to copper binding and transport (orange), carbohydrate binding and transport (violet), membrane binding and transport (grey) as well as other hypothetical proteins (white). The sequence data and annotation were taken from the Integrated Microbial Genomes database (IMG) (Markowitz *et al.* 2012).

3.7.3 Phylogenetic comparison of the newfound multicopper oxidases

The molecular phylogenetic analysis of all novel putative MCOs identified in this thesis was done by using the multiple sequence alignment tool M-coffee (Notredame *et al.* 2000; Wallace *et al.* 2006) by Maximum Likelihood method on the JTT matrix-based model. The tree with the highest log likelihood (Figure 29) was drawn to scale with branch length measured in number of substitutes per site.

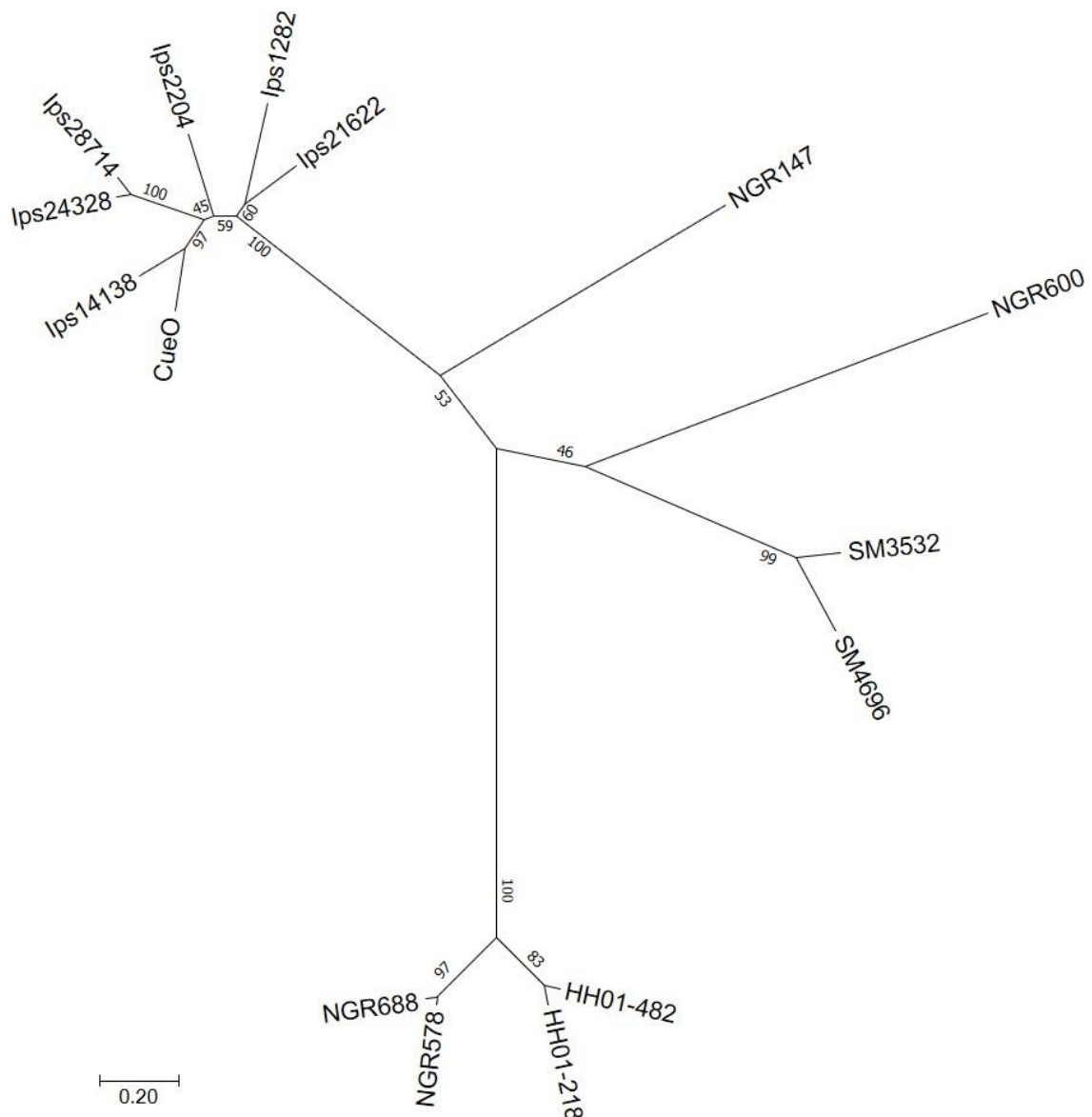


Figure 29: Molecular phylogenetic analysis of the 14 novel MCOs identified in this study in relation to CueO from *E. coli*. The evolutionary history of the proteins was analyzed per multiple sequence alignment tool M-coffee (Notredame *et al.* 2000; Wallace *et al.* 2006)(Notredame *et al.* 2000; Wallace *et al.* 2006) by Maximum Likelihood method on the JTT matrix-based model. The tree with the highest log likelihood (-5987.12) was drawn to scale with branch length measured in number of substitutes per site. Next to the branches, the percentage of calculated trees with exactly this structure was shown. The evolutionary analyses were done in MEGA7 (Kumar *et al.* 2018). This phylogenetic tree was kindly provided by Pablo Pérez-García (Department of Microbiology and Biotechnology, University of Hamburg).

As expected, the proteins discovered from the *lps typographus* metagenome were clustered together with CueO due to their common enterobacterial origin. Their arrangement was identical to the phylogenetic tree of their DNA sequences with only small deviations in percentages of calculated trees with exactly this layout. CueO clustered together with *lps14138* while *lps28714* and *lps24328* as well as *lps1282* and *lps21622* were also grouped as pairs, respectively. *lps2204* occupied the only singular branch while *lps28714* and *lps24328* showed the smallest sequence variation to each other in the enterobacterial cluster. The largest distance in paired proteins was found between *lps1282* and *lps21622*.

Interestingly, the four proteins from *S. fredii* NGR234 were not exclusively clustered together like the *S. maltophilia* K279a and *Duganella* sp. HH01 MCOs. Only NGR578 and NGR688 were sited in very close proximity to each other, while NGR147 was located directly next to the enterobacterial branch but still at a large distance. NGR600 was very remotely grouped together with the *S. maltophilia* proteins.

3.8 Vectormodification

Because MCOs need to bind copper as co-factor to build their active conformity, the pET22b(+) expression vector's own c-terminal His-tag was not suitable for purification as it forms chelate complexes with metal ions. Here, a modified pET22b(+) vector with an added StrepII-tag and a stop codon upstream of the His-tag – pET22b::StrepII – was used (kindly provided by Evocatal, Monheim am Rhein, Germany).

Because some of the novel MCOs contained an *NdeI* recognition sequence, it was not viable to use this restriction site for cloning into the original pET22b::StrepII (Figure 30). It was necessary to substitute the plasmid's sequence $5'GTATTT3'$ directly upstream of *NdeI* for the *NheI* recognition sequence $5'GCTAGC3'$ by means of site-directed mutagenesis (SDM). For this method, a primer-pair was designed with an overlapping area on the 5' end and an extended region on the 3' end. The mutation site for *NheI* was added to the primer that bound to the target region.

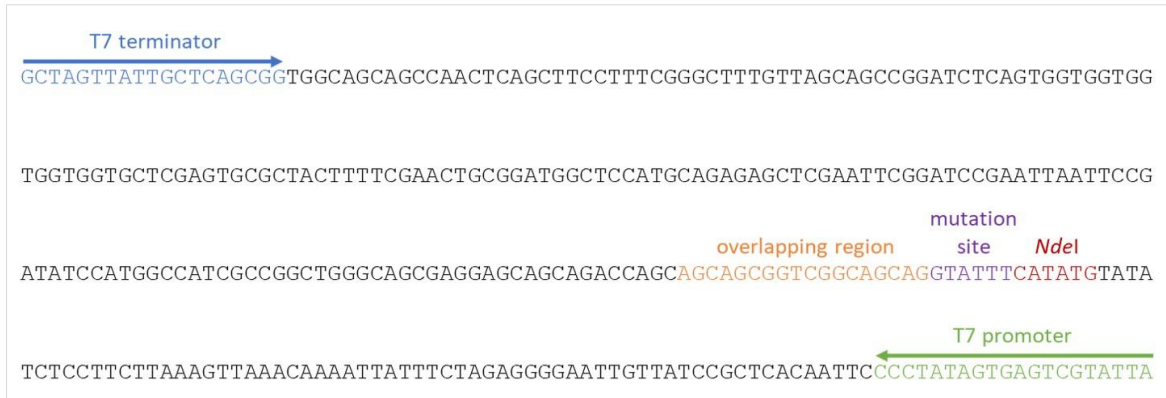


Figure 30: Original sequence of pET22b::StreptII in-between the T7 primer binding sites. The binding sites of the T7 primers were marked in blue (terminator) and green (promoter). The overlapping region of the SDM primers (orange) was located upstream of the mutation site (violet) and the *NdeI* recognition sequence.



Figure 31: Site-directed mutagenesis altered sequence of a clone originating from pET22b::StreptII. This clone exhibited three versions of the overlapping region with an additional copy of *NheI* and *NdeI* downstream of the third one. The sequence downstream of the second set of restriction endonuclease recognition sites was identical to its counterpart in the original vector sequence.

After SDM PCR, digestion with the endonuclease *DpnI*, and transformation into *E. coli* DH5 α , the T7 primers were used to check three different clones for sequence correctness.

One of these clones exhibited three copies of the overlapping region with *NheI* and *NdeI* sequences directly downstream of the first and third one (Figure 31).

For the next step, the clone's altered plasmid DNA was cut with *NheI*, purified of small fragments and religated to produce the target sequence of pET22b-SDM::StreptII with only one version of each endonuclease recognition sites. This newly modified vector was used to clone the MCO genes that contained an *NdeI* sequence themselves.

3.9 Amplification and cloning of the multicopper oxidase genes

After *in silico* identification of the previously introduced novel MCOs, the corresponding genes were amplified from their respective genomes or a sample of metagenomic *Ips typographus* DNA. For this purpose, the genomic DNA of *Duganella* sp. HH01, *S. maltophilia* K279a and *S. fredii* NGR234 was extracted and PCRs were carried out with specifically constructed primers (Table 8). It was not possible to amplify *sm4696* by PCR using the parameters tested in this study after being calculated by primer length and GC-content. Also, *ngr147* and *ngr688* were only successfully inserted into pDrive while the transfer into the expression vector was to no avail. The amplification of the bark beetle MCOs was executed with the same DNA sample that was used for the Illumina HiSeq sequencing of the *Ips typographus* dataset. It was managed to amplify *ips14138*, *ips21622* and *ips24328* from the residual metagenomic DNA. As the PCRs for *ips1282*, *ips2204* and *ips28714* were without success, the syntheses of these three gene sequences were ordered from Eurofins Genomics (Ebersberg, Germany). Prior to ordering, gene sequences and rare codons were optimized for expression in *E. coli* and restriction enzyme recognition sites were inserted on both ends. After production, the genes were shipped within pEX-K4 plasmids. Next, all genes were either intermediately cloned into pDrive and subsequently into pET22b::StreptII or pET22b-SDM::StreptII or they were inserted instantly into the expression vectors.

Both attempts, the direct amplification of *ips28714* from metagenomic DNA and the transfer of the synthesized gene into the expression vector were futile under applied experimental conditions. Overall, the construction of recombinant clones containing ten novel MCO genes was accomplished over the course of this thesis. All assembled MCO carrying expression plasmids were displayed and described in Table 7.

3.10 Heterologous gene expression

Eight of the ten gene constructs assembled in this study could be expressed heterologously in the *E. coli* hosts BL21(DE3) or T7 SHuffle by using auto-induction medium (Studier 2005). This method ensured a high cell density in the expression cultures prior to induction. The high glucose concentration in the medium suppressed promoter induction until $OD_{600} = 1$ was reached. Only afterwards, the metabolism of

lactose started and therefore mediated the induction of gene expression (Figure 32A).

BL21(DE3) is a standard expression strain containing the phage T7 RNA polymerase gene necessary for the function of T7 promoter carrying plasmids like pET vectors. *E. coli* T7 SHuffle offers the same features plus a constitutively expressed chromosomal copy of *dsbC*, which codes for a disulfide bond isomerase that acts as a chaperone and promotes the correct folding of enzymes with multiple disulfide bonds, like the tertiary structure of MCOs. As the identified metagenomic proteins were all derived from Enterobacteria, CueO from *E. coli* was consistently expressed additionally to check if the method itself was working properly for MCOs of related origin. The success of the expressions was verified by SDS-PAGE.

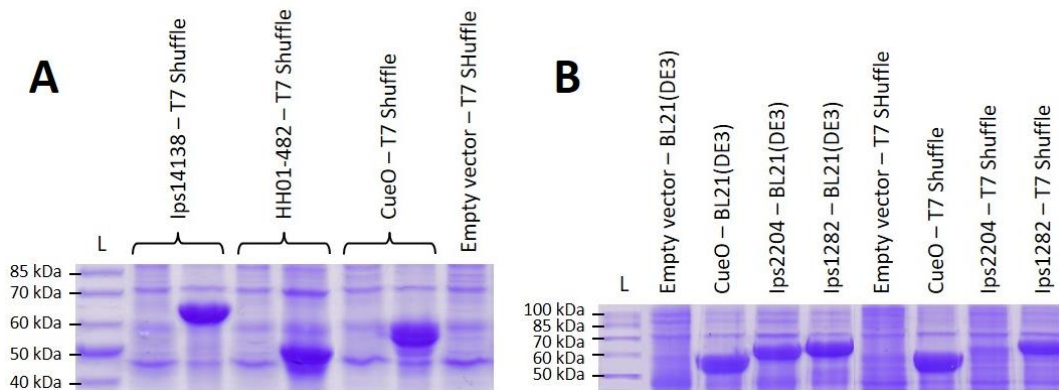


Figure 32: Expression of Ips14138 and HH01-482 in T7 SHuffle pre and past auto-induction (A) and comparison of protein yields of Ips1282 and Ips2204 in BL21(DE3) and T7 SHuffle, respectively (B). The first lane (L) in both gels contained the PageRuler™ Unstained Protein Ladder (Thermo Fisher Scientific, Braunschweig, Germany) and the expression strains carrying empty plasmids served as negative controls. The SDS gel **A** (12 %) showed the expression of Ips14138 (~62 kDa), HH01-482 (~48 kDa) and CueO (~55 kDa) at $OD_{600} = 0.6$ pre-induction (left lane) and after over-night expression (right lane), respectively. This demonstrated that there was no induction happening prior 0.6 OD_{600} and confirmed the working principle of the auto-induction method. SDS gel **B** (12 %) compared the Ips1282 (~65 kDa) and Ips2204 (~60 kDa) protein yields of both expression strains. It showed that Ips1282 was produced in large amounts by both strains while Ips2204 was only highly expressed in BL21(DE3). There was no considerable difference between the CueO concentrations produced by both strains.

It was not possible to express HH01-218 and NGR578 under the conditions tested in this thesis with neither BL21(DE3) nor T7 SHuffle. The amount of HH01-482 synthesized in T7 SHuffle (Figure 32A) was high and comparably lower in BL21(DE3). NGR600 showed the same distribution while SM3532 could only be produced in a very small quantity. Ips1282 yielded decent amounts of protein in both expression strains whereas Ips2204 was sufficiently expressed in BL21(DE3) and only slightly in T7 SHuffle (Figure 32B). Ips21622 and Ips24328 were both highly produced in

BL21(DE3) and to a low extent also in T7 SHuffle (Figure 33A). The highest overall quantities were yielded by *lps14138* in BL21(DE3) with an average of 20 mg/l protein (measured by Bradford protein assay) while the expression was also successful in T7 SHuffle.

The pelletized expression cultures that yielded a high amount of MCO were also visibly distinguishable from cells that only produced low concentrations or no protein at all. These cell pellets showed a greyish to blue-green color that correlated with the amount of active protein (Figure 33B). Based on soluble protein yield, for further experiments HH01-482 was expressed in *E. coli* T7 SHuffle and the *lps typographus* MCOs in BL21(DE3).

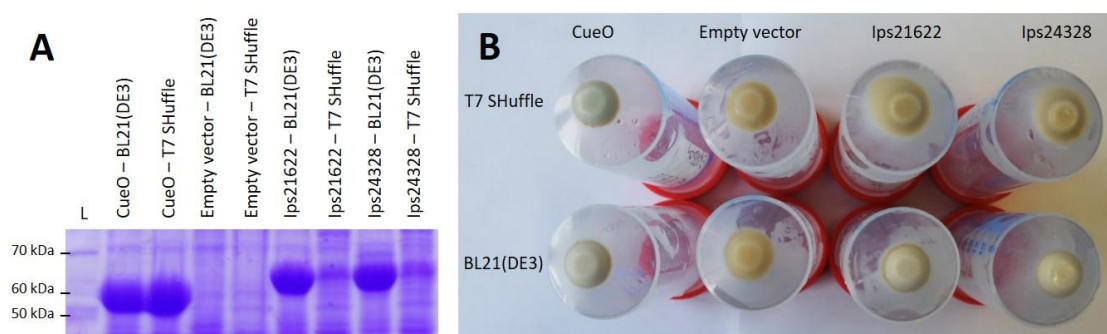


Figure 33: Expression of CueO from *E. coli*, *lps21622* and *lps24328* in BL21(DE3) and T7 SHuffle. The SDS gel **A** (12 %) showed the expression of CueO (~55 kDa), *lps21622* (~62 kDa) and *lps24328* (~63 kDa) in BL21(DE3) and T7 SHuffle, respectively. The first lane (L) contained the PageRuler™ Unstained Protein Ladder (Thermo Fisher Scientific, Braunschweig, Germany) and both expression strains carrying empty plasmid were used as negative controls. The gel showed that CueO was highly expressed in both host strains. *lps21622* and *lps24328* both yielded high amounts of protein in BL21(DE3) but only small quantities in T7 SHuffle. This correlated with the visual appearance of the cultures' pelletized cells (**B**). Both pellets of CueO showed a distinct bluish color that was slightly more intense for T7 SHuffle. The cells containing the empty vector were yellow-brown in color, as were the T7 SHuffle pellets of the metagenomic proteins. On the contrary, the BL21(DE3) cells producing *lps21622* and *lps24328* appeared a little lighter and more greyish than their counterparts.

3.11 Protein purification

After preparing clear cell lysate from the expression cultures, the enzymes were purified by Fast Protein Liquid Chromatography (FPLC) as described in chapter 2.7.3. Preliminary tests showed that PBS (pH 7.4) was better suited as binding buffer and elution buffer base than Tris-HCl. The mild pH also facilitated the protection of the proteins' tertiary and quaternary structure and therefore the enzyme activity. Following equilibration of the StrepTrap™ columns, the lysate was injected up

to five times to load the tagged protein on the resin. PBS was used to wash excess protein from the column and the elution was carried out with 2.5 mM desthiobiotin. It was possible to purify HH01-482, Ips1282, Ips2204, Ips14138, Ips21622 and Ips24328 over the course of this study. To exemplify the purification of these enzymes, the FPLC chromatogram of Ips1282 was displayed in Figure 34.

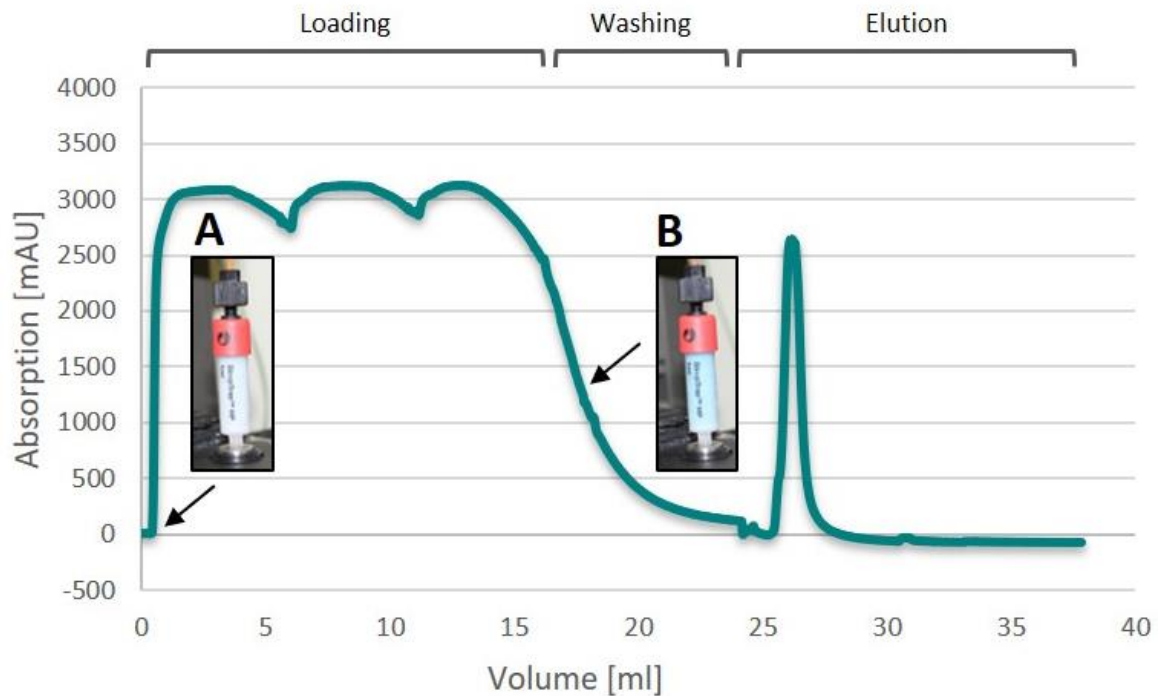


Figure 34: Purification of Ips1282 by Fast Protein Liquid Chromatography. Ips1282 was purified by affinity chromatography using the ÄKTAdesign™ FPLC system (Amersham Biosciences, San Francisco, USA) with a 1 ml StrepTrap™ HP column (GE Healthcare, Freiburg, Germany). The graph showed the absorption directly behind the column against the injected volume during loading, washing and elution of the protein. Initially, the resin was white (A) and turned blue-green after injection and binding of Ips1282 (B).

The purification of Ips1282 was initialized by loading the clear cell lysate on the column three times, which raised the absorption curve to about 3000 mAU and turned the resin blue-green. During the following washing step, the excess protein was flushed out until the absorption stopped decreasing further. The elution was carried out by substituting the bound protein with 2.5 mM desthiobiotin. The enlarged elution peak, corresponding SDS gel and image of the elution fractions were displayed in Figure 35. Ips1282 disassociated rapidly after induction of elution buffer under these experimental parameters. It already dissolved from the resin in the second elution fraction and showed the highest absorption value in fractions 3 and 4.

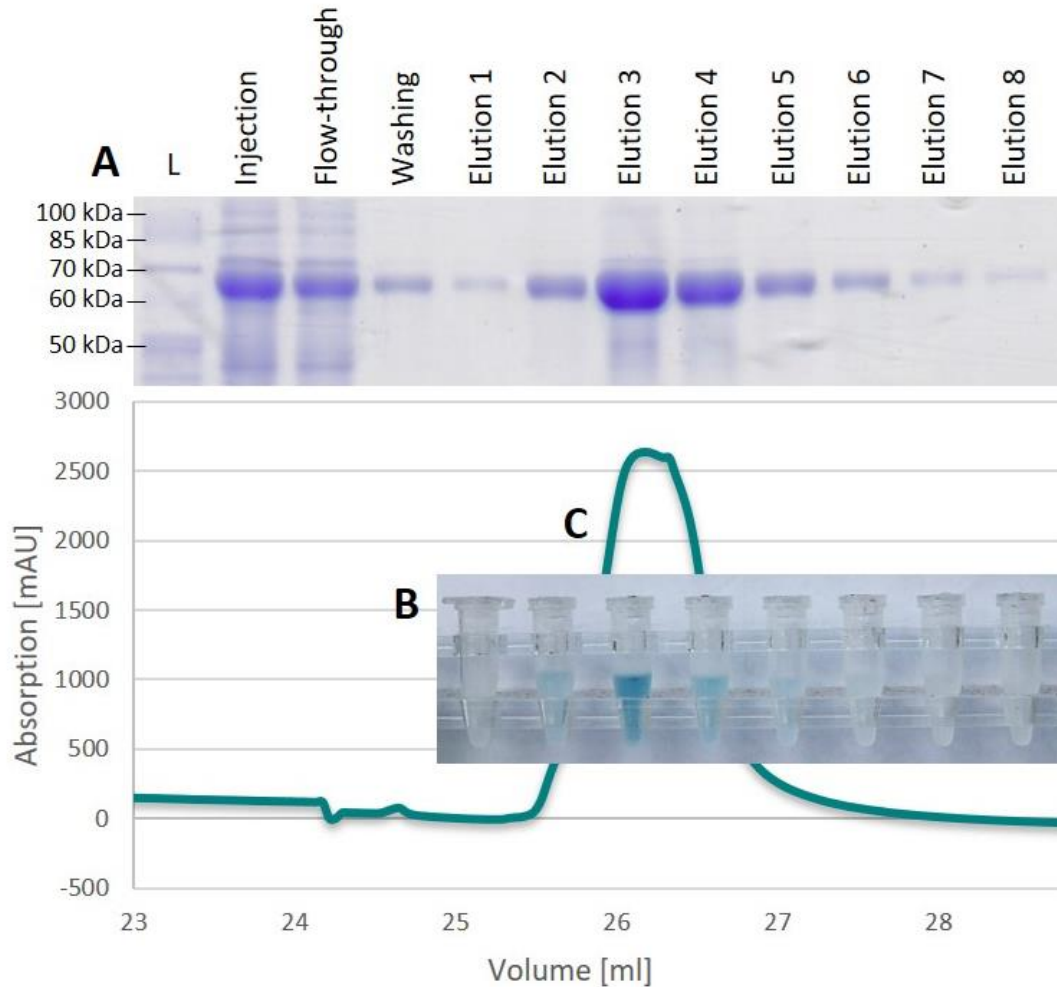


Figure 35: Elution peak of Ips1282 in correlating position to its SDS gel and protein fractions. The 12 % SDS gel (A) showed the purification steps of Ips1282 (~65 kDa) with injection, flow-through, washing and eight elution fractions. The left lane (L) contained the PageRuler™ Unstained Protein Ladder (Thermo Fisher Scientific, Braunschweig, Germany) for size comparison. The picture of the elution fractions (B) showed the blue-green color of the enzyme that correlated with the respective protein concentration and the peak of the FPLC curve (C) that displayed the highest absorption for elution steps 2, 3 and 4.

In general, the elution peak was comparably high and clear, indicating exceedingly pure protein which was confirmed by the corresponding SDS gel. The protein band ran at ~65 kDa while the amount of protein visual in each lane correlated precisely with the color intensity of the elution fractions. Overall, the purification of Ips1282 yielded approximately 3,5 mg of protein.

In comparison, Ips14138 demonstrated a similar color intensity after purification. Figure 36 visualized its gradual elution from the StrepTactin™ column evident by the migrating blue-green protein front.

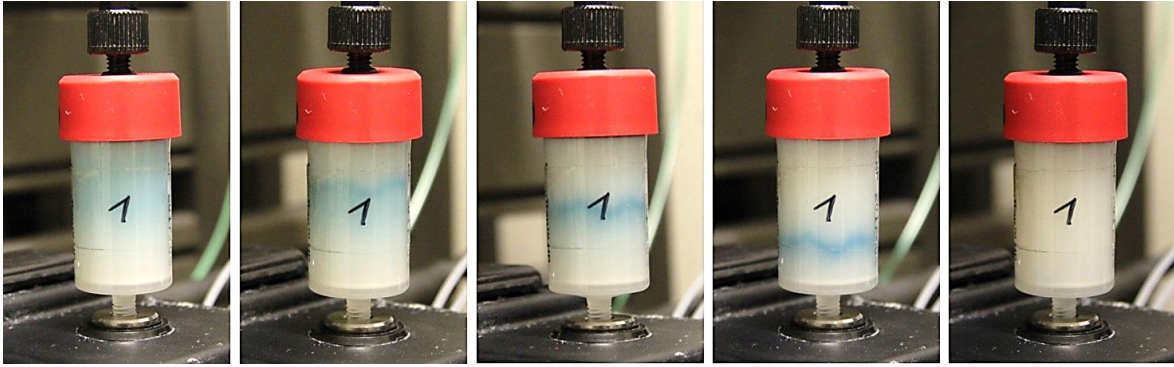


Figure 36: Gradual migration of the protein front during elution of Ips14138. All five pictures showed the 5 ml StrepTrap™ HP column (GE Healthcare, Freiburg, Germany) before and during elution of Ips14138. In the second to fourth image, the protein front migrated through the resin downwards before flushing out. The column was used in combination with the ÄKTAdesign™ FPLC system (Amersham Biosciences, San Francisco, USA).

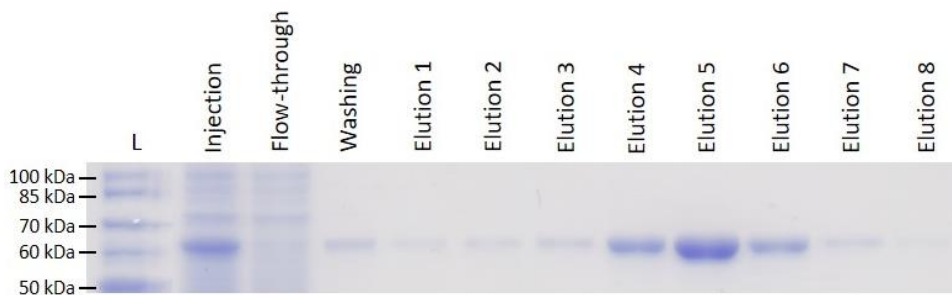


Figure 37: SDS gel of the purification of Ips14138. The SDS gel (12 %) of the purification of Ips14138 showed that the protein ran at ~62kDa and dissolved from the column in elution fractions 4 to 6 under these conditions. The first lane (L) contained the PageRuler™ Unstained Protein Ladder (Thermo Fisher Scientific, Braunschweig, Germany) for size comparison.

The SDS gel of the purification (Figure 37) showed that Ips14138 runs at ~62 kDa. It dissolved from the resin in elution fractions 4 to 6. This difference to Ips1282 can be explained by the use of a larger column which prevents a direct comparability. Ips14138 was the MCO that yielded the highest amount of protein with up to 25 mg per liter of expression culture. Comparative analyses showed no effect of desthiobiotin on the protein quantification.

In total, of all 14 novel putative MCOs identified during this study, 10 were successfully cloned into expression vectors (Table 40) while *ngr147*, *ngr688*, *sm4696* and *ips28714* were not. It was accomplished to express all of these in either *E. coli* BL21(DE3) or T7 SHuffle, except for *hh01-218* and *ngr578*. As Sm3532 was merely produced in an extremely small amount and NGR600 only yielded insoluble protein, it was not possible to purify them from crude extract. The six remaining enzymes

showed varying activity on ABTS at pH 3.6 and room temperature, resulting in further characterization of the four most promising candidates: HH01-482, Ips2204, Ips14138 and Ips24328.

Table 40: Comparison of experimental steps achieved for each novel putative MCO gene, with further characterized enzymes highlighted in bold.

Putative MCO gene	Cloning	Expression		Purification by FPLC	Activity on ABTS	Characterization
		BL21(DE3)	T7 SHuffle			
<i>ips1282</i>	+	+	+	+	±	–
<i>ips2204</i>	+	+	±	+	+	+
<i>ips14138</i>	+	++	+	+	+	+
<i>ips21622</i>	+	++	+	+	±	–
<i>ips24328</i>	+	++	+	+	+	+
<i>ips28714</i>	–	–	–	–	–	–
<i>hh01-218</i>	+	–	–	–	–	–
<i>hh01-482</i>	+	+	++	+	+	+
<i>sm3532</i>	+	±	±	–	–	–
<i>sm4696</i>	–	–	–	–	–	–
<i>ngr147</i>	–	–	–	–	–	–
<i>ngr578</i>	+	–	–	–	–	–
<i>ngr600</i>	+	+	++	–	–	–
<i>ngr688</i>	–	–	–	–	–	–
Total	10		8	6	6	4

Rating: – none/negative; ± very little; + positive; ++ strongly positive

3.12 Enzyme characterization

3.12.1 Structural properties

The four novel MCOs HH01-482, Ips2202, Ips14138 and Ips24328 were classified in group, superfamily (SF) and homologous family (HF) by using the Laccase and Multicopper Oxidase Engineering Database version 6.4 (LccED) (Sirim *et al.* 2011). The results in Table 41 showed that HH01-482 consisted of two domains and belonged to SF#14 (N - Bacterial type B 2dMCO) and HF#61. To date, there were no protein structures published for SF#14. There were however 21 structures for 2dMCOs collected in the Protein Data Bank (PDB) (Berman *et al.* 2000) of which 14

Table 41: Classification of the novel MCO protein sequences according to LccED*

Putative MCO	Genus / Best AA BLAST hit	Group	Superfamily	Homologous family
HH01-482	<i>Duganella</i> sp. HH01	2dMCO	SF#14	HF#61
lps2204	<i>Erwinia mallotivora</i> (78 %)	3dMCO	SF#10	HF#35
lps14138	<i>Cedecea neteri</i> (88 %)	3dMCO	SF#10	HF#35
lps24328	<i>Rahnella aquatilis</i> (94 %)	3dMCO	SF#10	HF#35

*<https://lcced.biocatnet.de/> access 2019-01-13 (Altschul *et al.* 1990; Sirim *et al.* 2011; Buchholz *et al.* 2016)

belonged to SF#11 (K - SLAC-like type B 2dMCO) and 7 were classified as SF#15 (O - Archaeal and Bacterial type C 2dMCO). All these previously known 2dMCO structures formed homotrimers which indicated that HH01-482 does as well (LccED access 2019-01-13).

To test this theory, the SWISS-MODEL workspace (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018) was used to predict the most probable 3D structure of HH01-482 (Figure 38). The software determined that mgLAC (PDB ID: 2zwn.1.A) (Komori *et al.* 2009) showed the highest sequence identity (30.5 %) to HH01-482 of all structures stored in the SWISS-MODEL Template Library (SMTL version 2018-11-14, PDB release 2018-11-09).

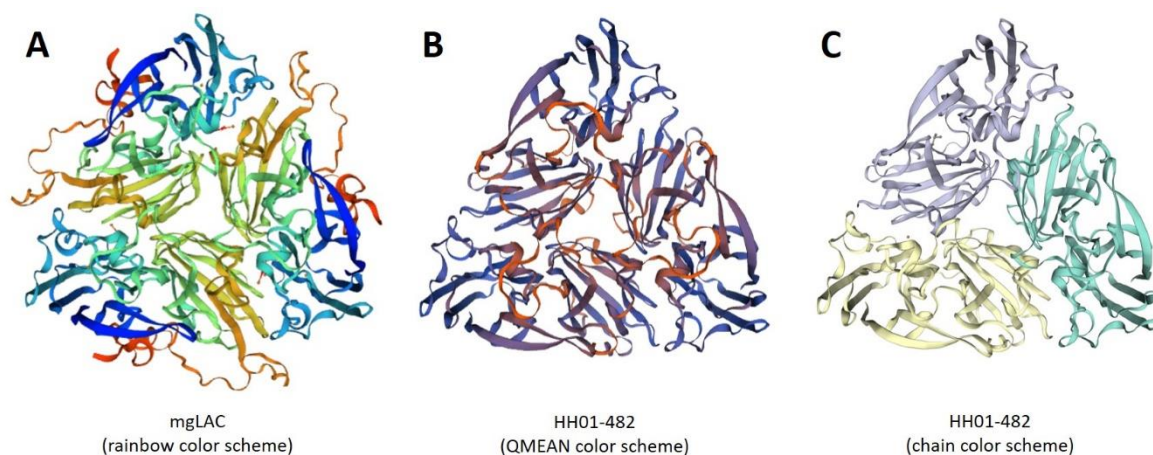


Figure 38: Template and model of the potential 3D structure of HH01-482. The potential protein structure of HH01-482 was modeled by submitting the sequence to the SWISS-MODEL workspace (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018) and illustrated by the NGL viewer (Rose *et al.* 2018). The template after which it was generated (A), was the metagenomic two-domain type laccase mgLAC (PDB ID: 2zwn.1.A; (Komori *et al.* 2009) that showed the highest identity (30.5 %) of all structures stored in the SWISS-MODEL Template Library (SMTL version 2018-11-14, PDB release 2018-11-09). The first trimeric HH01-482 structure (B) was displayed in QMEAN color scheme highlighting areas differing from the template structure in red. The second model (C) showed the three separate protein chains that formed the homotrimer quaternary structure of HH01-482.

This two-domain type laccase belonged to SF#15 and was metagenomic in origin. Following the 3D structure of mgLAC, SWISS-MODEL computed the most probable structure of HH01-482 to also be a homotrimer. The model slightly differed from the template, which was visualized in QMEAN color scheme that highlighted these areas in red (Figure 38B).

A second image of the quaternary structure of HH01-482 in chain color scheme emphasized the three monomers that were incorporated in the active protein form (Figure 38C). The rainbow color scheme of the template (Figure 38A) visualized the three separate protein chains as well by gradually changing color from N- to C-terminus. These illustrations showed that both trimeric proteins formed a triangle structure that contained another triangle in the center comprised of three beta sheet structures.

Overall, the triangular shape of the proteins was mostly identical, including the beta sheet structures in the three corners visualized in dark blue and the second yellow triangle in the center of the template. The loop structures located in the front of HH01-482 displayed the largest difference to the template protein. Furthermore, the template's three orange colored peripheral chains emanating from the corners of the inner triangle were absent in HH01-482. A calculation by the SignalP 4.0 software tool (Petersen *et al.* 2011) also revealed a signal peptide cleavage site between AA26 and 27.

The classification of the three characterized *Ips typographus* MCOs by LccED (Table 41) showed that they all contained 3 domains and that their best AA BLAST hit belonged to SF#10 (J - Bacterial CueO) and HF#35, exactly like CueO from *E. coli*. Ips2204 demonstrated the highest AA sequence similarity to a protein from *Erwinia mallotivora* (78 %) while Ips14138 correlated most with an enzyme from *Cedecea neteri* (88 %). The closest resemblance to already known MCO showed Ips24328 with 94 % sequence similarity to a protein from *Rahnella aquatilis*.

By submission of the three *Ips typographus* MCO sequences to the SWISS-MODEL workspace (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018), the most probable 3D structures were predicted. CueO (PDB ID: 5b7e.1.A) (Aker *et al.* 2016) showed the highest sequence identity of all structures stored in the SWISS-MODEL Template Library (SMTL version 2018-11-14, PDB release 2018-11-09) and was used as template to model the potential monomeric tertiary structures of Ips2204, Ips14138 and Ips24328 (Figure 39).

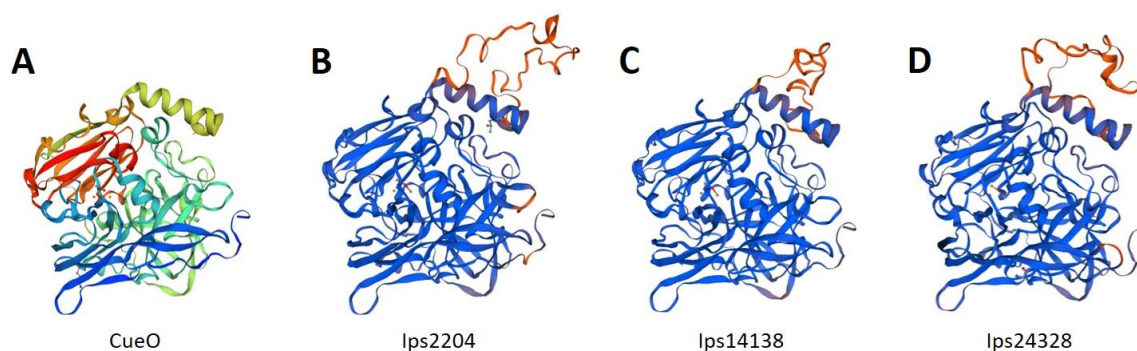


Figure 39: Template and models of the tertiary structures of Ips2204, Ips14138 and Ips24328. The potential protein structures of Ips2204 (B), Ips14138 (C) and Ips24328 (D) were modeled by submitting the sequences to the SWISS-MODEL workspace (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018) and illustrated by the NGL viewer (Rose *et al.* 2018). The template after which they were generated (A) was the crystal structure of deuterated CueO (PDB ID: 5b7e.1.A) (Akter *et al.* 2016). It showed the highest sequence identity of all structures stored in the SWISS-MODEL Template Library (SMTL version 2018-11-14, PDB release 2018-11-09) with 66.6 % to Ips2204, 75 % to Ips14138 and 64.3 % to Ips24328 with a coverage of 0.95, 0.98 % and 0.97 %, respectively. The models of the *Ips typographus* MCOs were displayed in QMEAN color scheme highlighting areas differing from the template structure in red.

The monomeric crystal structure of CueO (Figure 39A) was displayed in rainbow color scheme. The protein contained a series of prominent helical and beta sheet structures with the most distinct one – the methionine-rich helix – displayed in yellow on the upper right side of the image.

All three *Ips typographus* MCOs were illustrated in QMEAN color scheme, highlighting the areas differing from the template in red. Noticeably, the loop structure above the methionine-rich helix represented the greatest dissimilarity between the proteins. The template CueO did not possess this feature at all, while all three novel enzymes did. Aside from that, the red color was only evident in a few differing spots and in the lower right area, indicating slight sequence deviations for all three enzymes.

3.12.2 Storage stability

Preliminary enzyme activity tests on the three MCO substrates ABTS, syringaldazine and guaiacol (c.f. appendix Figure 51) showed the highest observed sensitivity for ABTS, whose amount of oxygenized product was measured at 420 nm. Based on these results, the enzyme activity analyses were carried out solely with using ABTS as substrate.

To investigate the stability of the four MCOs under different temperature and glycerol conditions, the activity on ABTS was measured before and after storage for 7,

14, 21 and 28 days, respectively. The temperature conditions included storage at 4 °C, 22 °C, 37 °C, and -20 °C with 0 %, 25 % and 50 % of glycerol, each. It was also determined, if using one and the same tube, refreezing, and storing it at -20 °C influenced the enzyme activity in contrast to using another aliquot for every measurement.

The comparison of the stability of the four enzymes (Ips2204 – Figure 40; Ips14138 – Figure 41; Ips24328 – Figure 42 and HH01-482 – Figure 43) showed that storage at -20 °C with 50 % glycerol was the most beneficial for the residual activity. All enzymes retained above 80 % of their initial activity after 28 days under these conditions. Ips14138 got even more active, rising to 115 % after 7 days and still stayed at approximately 105 % activity after 28 days of storage. A glycerol concentration of 25 % also kept the enzymes' activities above 80 % except for Ips24328, which fell slightly under 60 % after 7 days of storage and rose again to 68 % after 21 days.

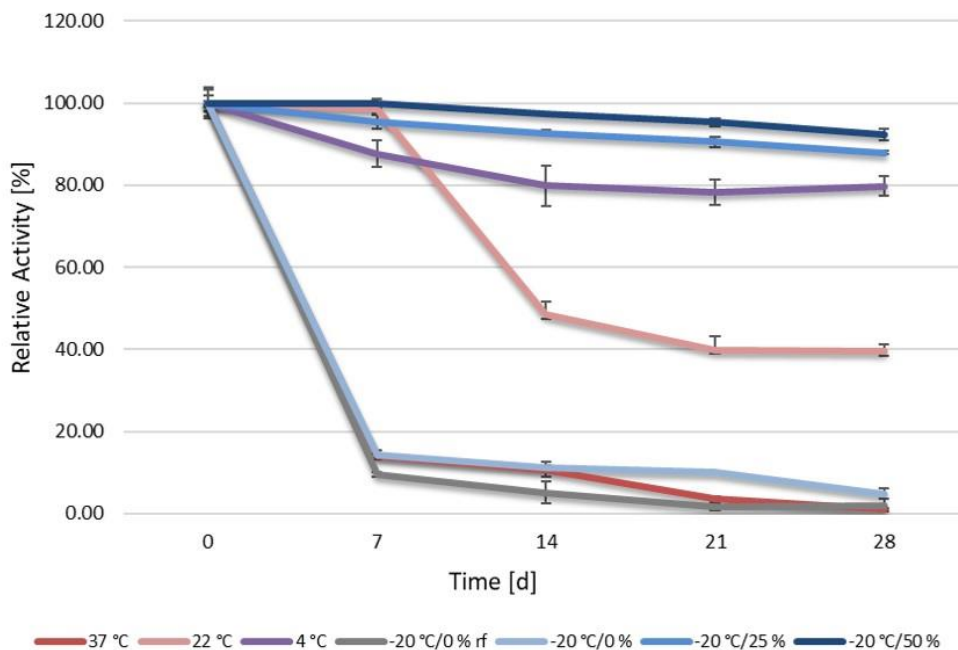


Figure 40: Storage stability of Ips2204 under different temperature and glycerol conditions.

The activity of Ips2204 on ABTS was determined immediately before and after storage in PBS buffer (7.4 pH) for 7, 14, 21 and 28 days at -20 °C, 4 °C, 22 °C and 37 °C, respectively. The difference between using one tube and refreezing it after each measurement (-20 °C/0 % rf) and using a new one every time, was also assessed as well as the effect of 0 %, 25 % and 50 % of glycerol on the enzyme activity after storage at -20 °C. The assay was carried out by incubating the reaction mixture for 20 min at room temperature and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

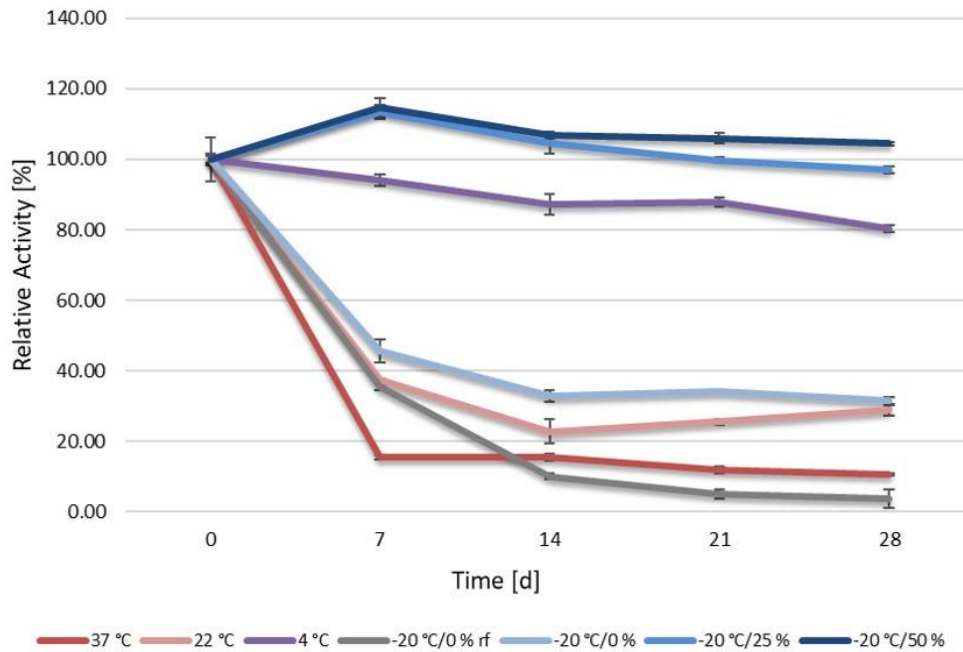


Figure 41: Storage stability of Ips14138 under different temperature and glycerol conditions. The activity of Ips14138 on ABTS was determined immediately before and after storage in PBS buffer (7.4 pH) for 7, 14, 21 and 28 days at -20 °C, 4 °C, 22 °C and 37 °C, respectively. The difference between using one tube and refreezing it after each measurement (-20 °C/0 % rf) and using a new one every time, was also assessed as well as the effect of 0 %, 25 % and 50 % of glycerol on the enzyme activity after storage at -20 °C. The assay was carried out by incubating the reaction mixture for 20 min at room temperature and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

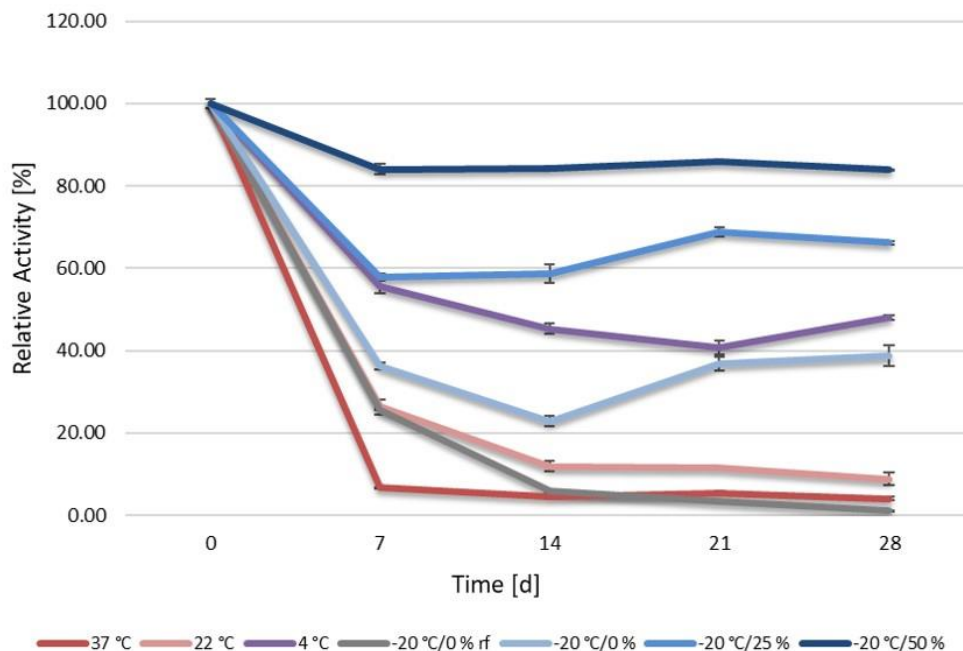


Figure 42: Storage stability of Ips24328 under different temperature and glycerol conditions. The activity of Ips24328 on ABTS was determined immediately before and after storage in PBS buffer (7.4 pH) for 7, 14, 21 and 28 days at -20 °C, 4 °C, 22 °C and 37 °C, respectively. The difference between using one tube and refreezing it after each measurement (-20 °C/0 % rf) and using a new one every time, was also assessed as well as the effect of 0 %, 25 % and 50 % of glycerol on the enzyme activity after storage at -20 °C. The assay was carried out by incubating the reaction mixture for 20 min at room temperature and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

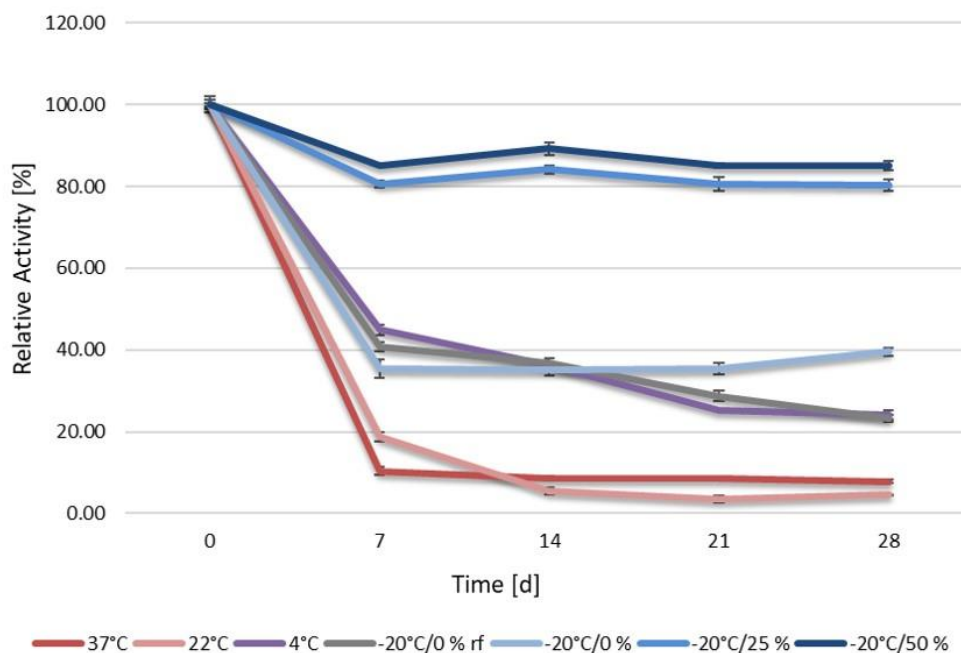


Figure 43: Storage stability of HH01-482 under different temperature and glycerol conditions.

The activity of HH01-482 on ABTS was determined immediately before and after storage in PBS buffer (7.4 pH) for 7, 14, 21 and 28 days at -20 °C, 4 °C, 22 °C and 37 °C, respectively. The difference between using one tube and refreezing it after each measurement (-20 °C/0 % rf) and using a new one every time, was also assessed as well as the effect of 0 %, 25 % and 50 % of glycerol on the enzyme activity after storage at -20 °C. The assay was carried out by incubating the reaction mixture for 20 min at room temperature and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

Ips2204 and Ips14138 showed similarities in the storage at 4 °C and -20 °C with 25 and 50 % glycerol as the values of all three slopes remained above 80 % after 28 days. Both also had 30 to 40 % of activity left after storage at 22 °C. Ips14138 showed nearly the same value after 28 days at -20 °C without glycerol as at 22 °C, whereas Ips2204 presented the lowest activity with <5 % for the samples stored at -20 °C without glycerol as well as at 37 °C. Ips24328 and HH01-482 exhibited values below 10 % after staying at 22 °C and around 40 % after -20 °C without glycerol for 28 days. Storage at 37 °C for 7 days reduced the residual activity of all four enzymes to <20 %. Ips2204 and Ips24328 fell below 5 % after 28 days at 37 °C while Ips14138 and HH01-482 still had 10 and 8 % activity left, respectively. Storage at 4 °C during the whole experiment left Ips24328 at 48 % and HH01-482 at 24 % residual activity. Except for Ips2204, there was a clear difference between storage at -20 °C with or without refreezing the tubes. All three *Ips typographus* derived MCOs presented no apparent residual activity after refreezing the tube following each measurement. In contrast, HH01-482 remained at 23 % despite refreezing while the remaining activity after 28 days at -20 °C stayed at 40 %. Ips14138 and

lps24328 presented even larger gaps due to refreezing with 28 and 37.5 % difference in activity, respectively. It was also noticeable, that *lps2204* still had about 40 % of its initial activity while *lps14138* fell to 30 % after 28 days at 22 °C.

3.12.3 Optimum pH

All four MCOs were analyzed by using a broad range of different pH values in the assay to examine the ideal pH for their activity on ABTS. The buffers for the reaction mixtures varied between pH 2.5 and 6.8 divided into citrate-phosphate (pH 2.5 to 3.2), acetate (pH 3.6 to 5.6) and phosphate buffers (pH 6.0 to 6.8) (cf. chapter 2.2). Comparing the four enzymes (Figure 44), *lps2204* and *lps14138* showed substantial similarities in their pH characteristics. Both exhibited the strongest activity on ABTS at pH 4.4 while *lps14138* demonstrated a broader pH range with above 40 % of its maximum activity between pH 2.8 and 5.6. The ideal buffer for *lps24328* possesses a pH of 3.6 while HH01-482 differed greatly from the metagenomic enzymes with a pH optimum of 5.6 while both show the same range of approximately 2 pH points above 20 % of their maximum activity.

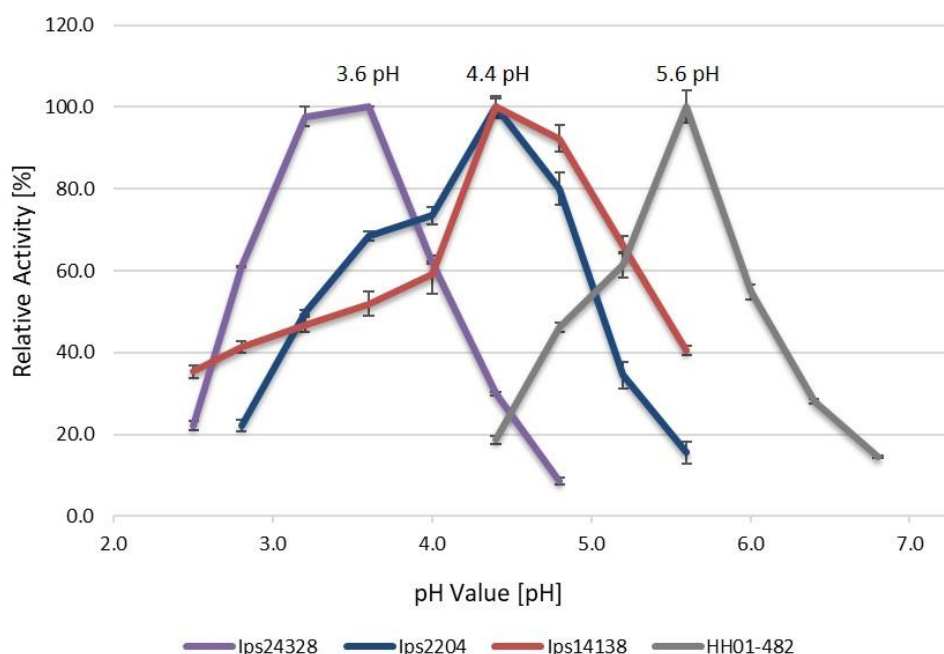


Figure 44: Optimum pH for the activity of *lps2204*, *lps14138*, *lps24328* and HH01-482 on ABTS. The pH optimum for the reaction of each enzyme on ABTS was determined using a variety of different buffers in the reaction mixture: citrate-phosphate buffer (pH 2.5, 2.8 and 3.2), acetate buffer (pH 3.6, 4.0, 4.4, 4.8, 5.2 and 5.6) and phosphate buffer (pH 6.0, 6.4 and 6.8). The assay was carried out by incubating the respective reaction mixtures for 20 min at room temperature and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

3.12.4 Optimum temperature

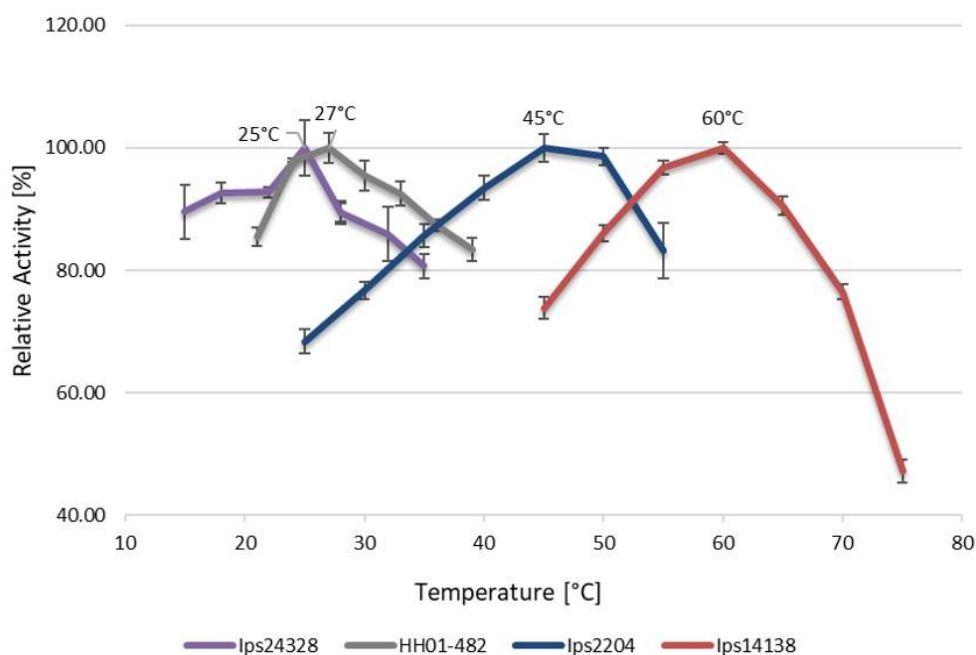


Figure 45: Optimum temperature for the activity of Ips2204, Ips14138, Ips24328 and HH01-482 on ABTS. The temperature optimum for the reaction of each enzyme on ABTS was determined by incubating the respective reaction mixtures for 20 min at a broad temperature range. The absorption of the oxygenized substrate was subsequently measured at 420 nm (n=3).

While Ips24328 and HH01-482 were most active on ABTS at 25 °C and 27 °C, respectively, for Ips2204 and Ips14138 higher temperatures were more beneficial (Figure 45). The tests showed an optimum temperature of approximately 45 °C for Ips2204 whereas Ips14138 presented the highest activity at 60 °C.

3.12.5 Heat stability

The heat stability of the four MCOs was measured by pre-incubating freshly prepared protein at 95, 80 and 65 °C for different lengths of time. The respective reaction mixtures featured a beneficial pH for each enzyme (Ips14138 – 4.5 pH; HH01-482 – 5.6 pH; Ips24328 – 3.6 pH and Ips2204 – 4.5 pH) while the assay was executed at 25 °C. To eliminate inaccuracies due to a delay in reaching the appropriate temperature, the buffers were pre-heated prior to enzyme dilution.

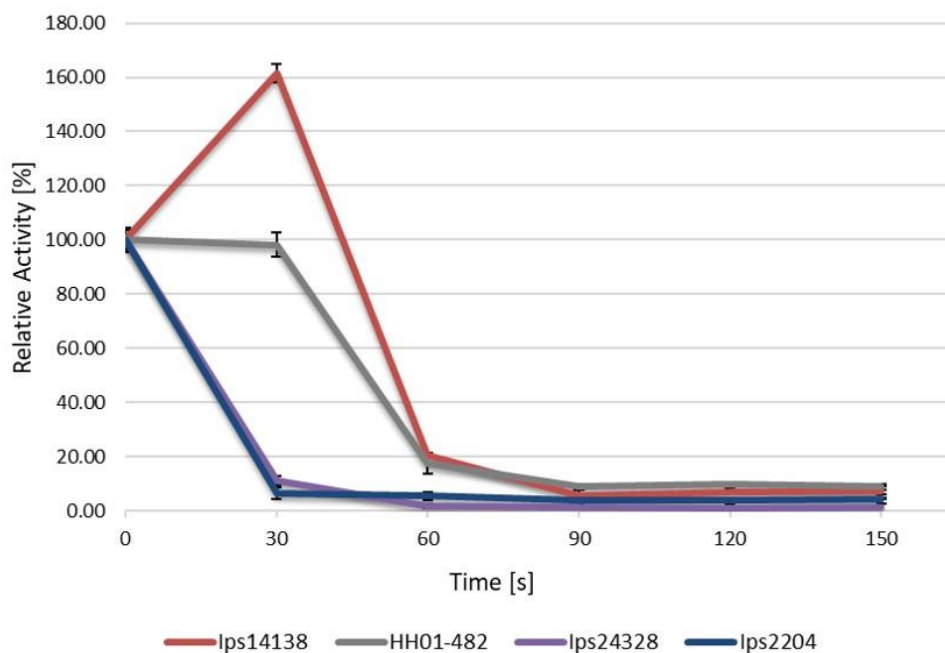


Figure 46: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 95 °C. The activity on ABTS was measured after incubating the enzyme solutions for 30, 60, 90, 120 and 150 s at 95 °C. The PBS (pH 7.4) used to dilute the enzymes was pre-heated to 95 °C. Individual buffers with a beneficial pH for each enzyme were used for the reaction mixtures: Ips2204 - pH 4.5, Ips14138 - pH 4.5, Ips24328 - pH 3.6 and HH01-482 - pH 5.6. The assay was carried out by incubating for 20 min at 25 °C and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

During incubation at 95 °C (Figure 46), the activity of Ips24328 and Ips2204 already dropped to about 10 % and lower after 30 s. At the same time, HH01-482 stayed equal to the starting value while decreasing to 20 % between 30 and 60 s before evening out to approximately 10 % after 90 s. In contrast, Ips14138 rose to a relative activity of 160 % after 30 s before steeply dropping to 20 % after 60 s and evening out similar to HH01-482.

The incubation at 80 °C (Figure 47) showed that Ips2204 was the least heat-stable of the four enzymes under these conditions. After approximately 240 s of impact, it already reached its lowest relative activity of <5 %. Ips24328 and HH01-482 were slightly more stable as they reached their lowest activity after about 540 s. In contrast to Ips24328 which declined steadily, HH01-482 rose to 130 % after 30 s and then decreased more steeply until it reached percentages similar to Ips24328 again after 120 s. The positive effect of high temperatures on the initial activity of Ips14138 was even higher at 80 °C. In comparison to the enzyme activity without heat impact, the slope rose to almost 250 % over the course of 90 s and remained above 100 % until 300 s passed. After approximately 580 s, the enzyme activity fell to <5 %.

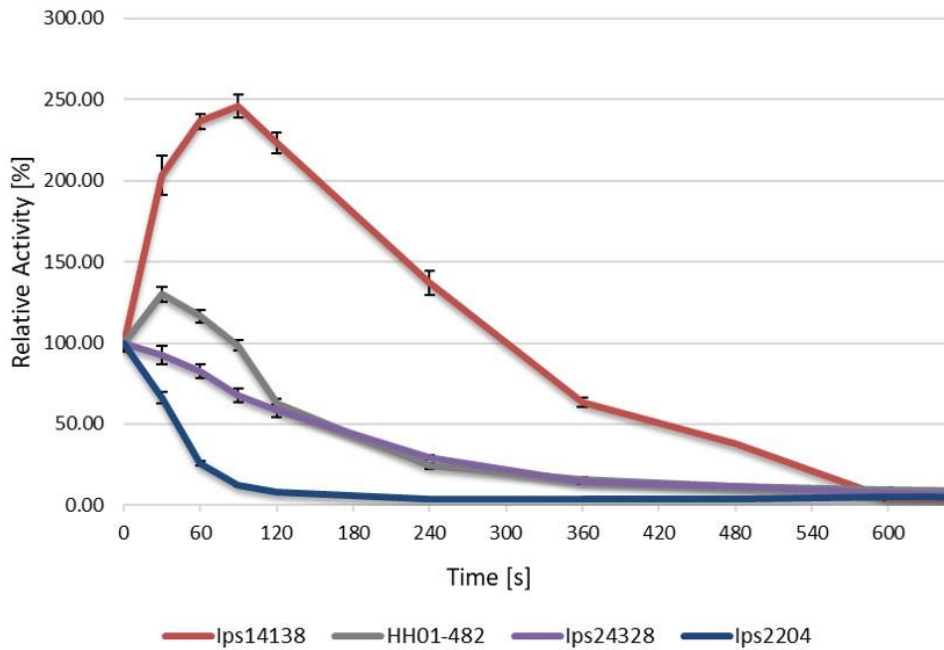


Figure 47: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 80 °C. The activity on ABTS was measured after incubating the enzyme solutions for up to 600 s at 80 °C. The PBS (pH 7.4) used to dilute the enzymes was pre-heated to 80 °C. Individual buffers with a beneficial pH for each enzyme were used for the reaction mixtures: Ips2204 - pH 4.5, Ips14138 - pH 4.5, Ips24328 - pH 3.6 and HH01-482 - pH 5.6. The assay was carried out by incubating for 20 min at 25 °C and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

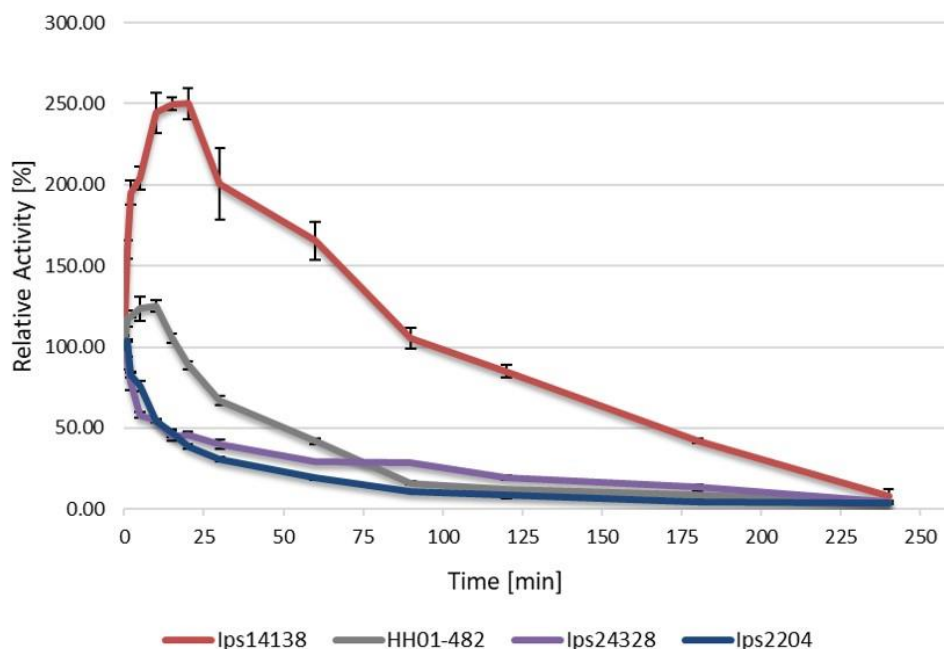


Figure 48: Stability of Ips2204, Ips14138, Ips24328 and HH01-482 during heat treatment at 65 °C. The activity on ABTS was measured after incubating the enzyme solutions for up to 240 min at 65 °C. The PBS (pH 7.4) used to dilute the enzymes was pre-heated to 65 °C. Individual buffers with a beneficial pH for each enzyme were used for the reaction mixtures: Ips2204 - pH 4.5, Ips14138 - pH 4.5, Ips24328 - pH 3.6 and HH01-482 - pH 5.6. The assay was carried out by incubating for 20 min at 25 °C and measuring the absorption of the oxygenized substrate at 420 nm (n=3).

The diagram of the heat stability test at 65 °C (Figure 48) resembled the previous assessment at 80 °C with the incubation time being the most apparent difference. Ips14138 again increased to about 250 % of the activity of non-heat-treated protein. Under these conditions, it reached the maximum activity after 20 min and fell back to 100 % after about 100 min of incubation. After 3 h at 65 °C Ips14138 still demonstrated more than 40 % of the non-heat-treated protein activity.

HH01-482 initially rose to almost 130 % of its unimpaired protein activity before it descended to <5 % after 240 min. Ips24328 and Ips2204 both decreased to under 50 % of their non-heat-treated activity after 15 min of incubation at 65 °C. Afterwards, the activity steadily declined until the end of the measurement after 4 h.

4 Discussion

4.1 Sample collection, DNA extraction and sequencing

Due to the small size of the specimens of 4.2 to 5.5 cm, several hundred beetles were needed to provide a sufficient quantity of intestines to proceed. The small size of the insects was also reason for the difficulty of separating beetle tissue and feces. Even though the lysis conditions of the DNA extraction kit supposedly increased the ratio of bacterial to eukaryotic DNA, the initial analysis of taxonomic distribution revealed almost 40 % Eukaryota. It is questionable if a different protocol would have impaired this ratio.

The 4.6 ng of isolated metagenomic DNA were slightly too low for the NEBNext[®] library preparation kit, as 5 ng were the manufacturer's recommended minimum (New England BioLabs, Frankfurt am Main, Germany). This posed a considerable risk for the success of this preliminary sequencing step but ended up being of no obvious negative consequence. The assembled *Ips typographus* metagenome counted 219 Mbp which appears to be a typical size as it was almost the average of the randomly chosen insect datasets ranging between 143 and 281 Mbp. If the metagenomic sequencing was done again today, it would need much less DNA to get more reads and consequently a more complete dataset due to the rapid technological advances in this field.

4.2 Taxonomic distribution and genome size

The taxonomic hits distribution of the whole *Ips typographus* metagenome showed that 59.7 % of the data originated from bacteria, equating about 131 Mbp of sequences. Distributed into the predicted 140 OTUs, this resulted in an average length of 936 Kbp per bacterial genome. This correlated with data from literature, as McCutcheon and Moran declared the size of bacterial genomes covering from about 4.6 Mbp for free-living bacteria (e.g. *E. coli*) to 500 kbp and lower for symbiotic bacteria (McCutcheon & Moran 2011). An average of 936 kbp also pointed to several genomes being on the smaller side, suggesting a possible deletion of genes unnec-

essary for symbiotic lifestyle. Especially maternally transmitted or “secondary” symbionts exhibited genome reduction while these were reported to involve Enterobacteriaceae in aphids and often being related to pathogens (Burke & Moran 2011). As Proteobacteria, including pathogens like *Escherichia* and *Yersinia*, contributed almost 50 % of the *Ips typographus* sequences (Figure 17), there is a large possibility that genome reduction might also have happened in this habitat.

Of the eukaryotic groups found by taxonomic hits distribution, Chordata, Echinodermata, Cnidaria and Hemichordata seemed to be falsely assigned or due to sample contamination. Some Nematodes are known to have a phoretic or parasitic association with insects, explaining the occurrence of 0.60 % correlated sequences (Holterman *et al.* 2017). The discovery of 0.16 % Microsporidia was likely attributed to the presence of *Chytridiopsis typographi*, a microsporidian pathogen of *Ips typographus* previously recovered from bark beetle specimens in the Czech Republic and Poland (Tonka & Weiser 2010; Corsaro *et al.* 2019). Streptophytal sequences with 0.12 % of the total dataset were possibly caused by plant material ingested during feeding, leaving a remnant of 0.22 % unclassified eukaryotes and viruses.

4.3 Phylogenetic analysis by 16S rRNA gene sequencing

The evaluation of potential primer pairs prior to 16S rRNA gene amplification gave the advantage of reducing the risk of disregarding genera typically present in this kind of environment. Without previous *in silico* analysis, performance and coverage may have been less appropriate due to primer bias.

Typically, the diversity of gut microbial communities of wood-feeding insects appears to depend on range of host trees. Polyphagous representatives that can colonize many different host trees like the Cerambycidae *Anaplophora glabripennis* tend to have more diverse communities varying by host species. Previous studies show an estimation of over 300 OTUs in association with *A. glabripennis* (Scully *et al.* 2013). As *Ips typographus* has a restricted host tree range – the principal host being the Norway spruce *Picea abies* – its gut microbial community should present comparably less complex and diverse.

The sequencing of 16S rRNA genes amplified from the *Ips typographus* metagenomic DNA resulted in 113 unique OTUs while predicting the rarefaction satura-

tion at about 140, representing less than half of the *A. glabripennis* microbial community. These sequences consisted of 91.15 % Proteobacteria, 7.96 % Firmicutes and 0.88 % Actinobacteria, also confirming the bacterial phyla identified by taxonomic hits distribution (Figure 17). In comparison, the distribution of all 1667 identified sequences showed 97.48 % Proteobacteria, 2.46 % Firmicutes and 0.06 % Actinobacteria. This indicates that additionally to contributing the largest number of OTUs, proteobacterial genera seem to have had a higher relative abundance in the gut bacterial community compared to the other phyla. Another possible explanation to consider is a preferential amplification of the 16S rRNA primers for Proteobacteria, especially *Morganella*.

Cross-taxon comparisons of microbial communities show that proteobacterial taxa are dominant in a large variety of different insect species, correlating perfectly with the results for *Ips typographus*. Also, the most abundant phylotype typically accounts for an average of 54.7 % of the whole community, being represented in this environment by *Morganella* with 57.1 % of all 16S rRNA sequences and 31.9 % of all OTUs. In general, insect-associated bacterial communities tend to harbor a limited bacterial diversity usually consisting of fewer than 8 phylotypes (Jones *et al.* 2013). Ignoring genera with less than 1 % of abundance, there were 7 different phylotypes present which supports this conclusion from literature.

Previous research on bark beetle associated microbial communities also showed limited diversity and a dominance of Proteobacteria, confirming the phylogenetic findings of this study. Larvae, pupa and imago of the pine engraver *Ips pini* accounted for 5 different genera comprising of γ -Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes (Delalibera *et al.* 2007). This also holds true for two *Dendroctonus* species, the southern pine beetle *D. frontalis* and the eastern larch beetle *D. simplex*. The first one yielded 5 different taxa from larval and 9 from adult specimens by 16S rRNA analysis, all of these derived from γ -Proteobacteria and Firmicutes (Vasanthakumar *et al.* 2006). Endomicrobial investigation of three *D. simplex* samples delivered 4 distinct Phylotypes from γ -Proteobacteria and Firmicutes in addition to 20 % to 35 % of sequences originating from non-abundant bacteria (< 1 %) (Durand *et al.* 2015). This also represents a similarity to the taxonomic distribution of the *Ips typographus* gut microbiome, which exhibited 15 % of non-abundant OTUs itself. Regarding the not fully saturated rarefaction curve, this percentage will likely rise with additional 16S rRNA reads.

4.4 Phylogenetic analysis by 18S rRNA gene sequencing

The 18S rRNA gene amplification on metagenomic DNA isolated from *Ips typographus* yielded 32 eukaryotic OTUs. Six relevant sequences were left after removing the ones from beetle and plant origin. Nematodes of the genus *Mycoletzkyia* delivered two OTUs, reinforcing the detection of equivalent sequences in the metagenomic dataset (Figure 17). These known bark beetle symbionts were previously described to solely inhabit Scolytinae and *Araucarius*, closely related wood-boring weavels. *Mycoletzkyia* appear to be highly host-specific as a previous study by Susoy and Herrmann identified 26 different putative species of which 19 were found in an association with a single bark beetle species (Susoy & Herrmann 2014). The two sequences found in this study's *Ips typographus* specimens showed the highest identity of 98.87 % and 99.25 % to *Mycoletzkyia buetschlii*, which was exclusively observed in the *Ips* genus, corroborating the results from literature.

One OTU was assigned to the oomycete *Phytophthora*, an important plant pathogen amongst others responsible for the notorious potato blight (*P. infestans*) or the sudden oak death (*P. ramorum*) (Yang *et al.* 2017). A BLAST search (access 2020-02-20) showed it was closely related to *P. alticola* and *P. frigida*, two species from South Africa that attack eucalyptus (Maseko *et al.* 2007), and the Dutch rose pathogen *P. bisheria* (Abad *et al.* 2008) with 97.63 % sequence identity, each. Previous research reveals that *Phytophthora* infection and canker formation attracts ambrosia and bark beetles to potential host trees with a single finding of *Ips* as well (Kelsey *et al.* 2013). Since there is no common knowledge of *Phytophthora* being insect gut associated, they could have been internalized by the beetle during feeding, although a phoretic relationship as well as the transition to an exclusively host-related lifestyle cannot be discounted.

There were two OTUs from Ascomycota and one from Basidiomycota identified by 18S rRNA sequencing. Fungi in general are frequently found in the guts of wood-feeding insects, playing a key role in digestion of sturdy plant material (Engel & Moran 2013). One of the ascomycotal OTUs most likely originated from *Wickerhamomyces bisporus* or a closely related yeast. These fungi are known to be insect-associated by example of the malaria vector *Anopheles stephensi* and *Drosophila* (Ricci *et al.* 2011; Stefanini 2018). *Wickerhamomyces mori* was also isolated from

unspecified wood-boring larvae – possibly Scolytidae – collected from a white mulberry tree (*Morus alba*) in central China (Hui *et al.* 2013). On these grounds, it can be assumed that the *Wickerhamomyces* OTU likely belongs to an intestinal symbiont of *Ips typographus*.

The second ascomycotal OTU found derived from an unknown Sarcosomataceae, which consist of fungi living purely saprotrophic on rotten wood, in endophytic symbiosis with plants or as their parasites (Köpcke *et al.* 2002). There is no known case in literature to date of Sarcosomataceae being insect gut symbionts, which points to the conclusion that the material was likely ingested during feeding. Same as with *Phytophthora*, a phoretic or yet unknown symbiotic relationship is still a possibility. The sole basidiomycotal sequence was identified as *Malassezia globosa*. Because this fungus is associated with human skin and frequently responsible for dandruff development and other skin disorders (White *et al.* 2014), its presence was likely due to contamination.

4.5 Annotation and comparison of different phylogenetic approaches

A comparison of the phylogenetic approaches (Table 42) showed that even though Proteobacteria contributed 97.48 % of all 16S rRNA sequences amplified from *Ips typographus* intestines, they only counted for 48.43 % of all protein-coding genes in the annotated metagenomic dataset. That either points to a significant reduction of genes in the enterobacterial genera as mentioned before or to a large number of genome copies in the dataset.

Table 42: Phyla comparison of the *Ips typographus* 16S rRNA sequences and all annotated protein-coding genes.

Bacterial phyla	16S rRNA OTUs	All 16S rRNA sequences	Protein-coding genes (IMG)
Proteobacteria	91.15 %	97.48 %	48.43 %
Firmicutes	7.96 %	2.46 %	22.85 %
Actinobacteria	0.88 %	0.06 %	12.73 %
Bacteroidetes	-	-	5.64 %
Cyanobacteria	-	-	2.40 %
Spirochaetes	-	-	1.85 %
Tenericutes	-	-	1.05 %

Phyla with less than 1 % of protein-coding genes were not shown.

Interestingly, IMG identified 5.64 % of all protein-coding genes as Bacteroidetes even though this phylum was not discovered by 16S analysis, same as Cyanobacteria, Spirochaetes and Tenericutes with 2.40 %, 1.85 % and 1.05 %, respectively. As Actinobacteria were assigned to 12.73 % of the protein-coding genes, equaling one single OTU hit, phyla with less than 10 % may not be abundant enough in the metagenomic template DNA to amplify even one sequence. These phyla may very well be part of the approximately 30 undetected OTUs the rarefaction slope indicated. Another reason could be underrepresented amplification due to primer bias.

4.6 Comparison with other gut-associated metagenomic datasets

In comparison to other host-associated metagenomic datasets, the *Ips typographus* gut microbiome showed the second highest rate of protein-coding genes related to Proteobacteria, Spirochaetes and Actinobacteria while it accounted for the second smallest number of genes from Firmicutes and Fusobacteria origin. This unequal distribution of phyla also stood out in the datasets of *Sirex noctilio* and *Ailuropoda melanoleuca*, which both possess a limited nutritional range. The woodwasps feed on pine wood previously modified and predigested by their symbiotic fungus *Amylostereum areolatum* (Lombardero *et al.* 2016) while the diet of the giant panda consists of 99 % bamboo (Sponheimer *et al.* 2019).

As *Ips typographus* specimens almost exclusively inhabit their main host tree *Picea abies*, their nutrition does not vary considerably. In contrast, the other two wood-feeding beetles *Anoplophora glabripennis* and *Viturius sinuatocollis* both feed on a wide variety of hardwood tree species (Urbina *et al.* 2013; Mason *et al.* 2016). Comparing the phylogenetic distribution of all three beetles, the ratio of Firmicutes to Proteobacteria was more equal for *A. glabripennis* and *V. sinuatocollis* than for *Ips typographus*. This stronger uniformity was also represented in the datasets of organisms with a more diverse nutrition like *Gallus gallus* and *Canis familiaris*.

Possibly, this uneven distribution of phyla is characteristic for organisms with a highly specialized food source together with the limited diversity in intestinal environments, that was described in introduction section 1.2 of this study.

Regarding functional comparison of the datasets, the bark beetle possessed the highest number of protein-coding genes classified in four different KEGG categories, all associated with varying metabolic activity. This clearly confirms the assumption,

that the gut of a highly specified wood-feeding beetle represents an extremely promising source for novel metabolic enzymes. The potential of the *Ips typographus* metagenome not only lies in the discovery of unknown MCOs, but evidently in metabolic fields like AA, carbohydrate, and lipid metabolism as well, waiting for exploration.

4.7 Novel putative multicopper oxidases

4.7.1 Enzymes originating from the *Ips typographus* dataset

All genera assigned to the six putative MCO genes (Table 38) except Rouxiella were also previously identified as part of the *Ips typographus* microbiome by 16S rRNA sequencing. It was noticeable, that each gene hit was matched to a different genus of Enterobacteria. This suggests that each genus only carries one version of this MCO, presumably a homologue to CueO since it clustered together with all identified metagenomic MCO genes in the phylogenetic tree (Figure 25). Interestingly, even though there were fungal protein-coding genes and 18S rRNA OTUs identified in the dataset, there were no viable eukaryotic MCO genes discovered.

Because of low-ranking query cover, sequence identity and therefore similarity to already known genes, *ips2204* and *ips14138* were the most promising candidates, followed closely by *ips1282*.

Regarding the phylogenetic analysis of the metagenomic genes (Figure 25), *ips21622* and *ips1282* were the most diverse sequences compared to *cueo*, defining them as interesting candidates as well. Because of the phylogenetic resemblance of *ips14138* to *cueo*, similar characteristics concerning temperature or pH performance could potentially be expected. In comparison to the other bacterial MCOs identified in this study (Figure 29), all enzymes derived from the *Ips typographus* metagenome clustered together with CueO due to their common enterobacterial origin. Their positioning and distance from each other were nearly identical in both phylogenetic trees.

Concerning the proteins' copper binding motifs, the most conspicuous aspect was the high sequence diversity of *Ips2204* in contrast to CueO and the other five metagenomic enzymes. Surrounding CBM one and three, *Ips2204* displayed unique AAs not located in any other protein sequences shown in this lineup (Figure 26). The only deviation directly inside the CBM was one isoleucine substituted by valine

in the sequence of Ips1282. Also, the positions of all CBMs of Ips1282 were consistently located the furthest downstream, increasing its overall potential.

CBM sequences of Ips24328 and Ips28714 were notably similar with only one visible variation in AA position 137. This reinforced the findings of the phylogenetic analysis, confirming the close relation of both enzymes to each other.

4.7.2 Enzymes originating from other bacterial strains

About the MCOs originating from accessible bacterial strains, it was noticeable that *S. fredii* NGR234 provided four diverse putative MCO enzymes while *S. maltophilia* and *Duganella* sp. HH01 only possessed two distinct sequences.

This feature may be connected to the ability of inhabiting an exceptionally broad range of host plants and leads to the assumption, that MCO activity may be involved in colonization processes. In literature, there was lignolytic activity previously discovered in *Sinorhizobium meliloti* – a close relative of *S. fredii* – and other Rhizobiaceae (Pawlik *et al.* 2016; Wang *et al.* 2016), hinting at considerable MCO potential in NGR234.

Comparing sequences belonging to and surrounding the four typical CBM, it was observed that NGR578 and NGR688 were completely identical and showed more resemblance to the HH01 proteins than to the other two also deriving from *S. fredii*. A reason for this could originate from their affiliation to the same SF and from possessing the same domain composition. This finding was confirmed by the phylogenetic comparison of all enzymes identified in this study (Figure 29).

In this phylogenetic tree, the four enzymes obtained from *S. fredii* NGR234 did not all cluster together like the candidates from *S. maltophilia* and *Duganella* sp. HH01. Only NGR688 and NGR578 were located directly next to each other while being on the same branch and in close proximity to the HH01 proteins. NGR147 was positioned directly next but still at a large distance to the enterobacterial branch and NGR600 was closer grouped with SM3532 and SM4696 than with any of the other proteins. This large variety of NGR234's MCOs may increase its substrate flexibility and therefore enhance its capability of populating a broad range of different plants. The inspection of ORF neighborhoods (Figure 28) also revealed a high similarity of regions surrounding *ngr578* and *ngr688*, mirroring prior findings. The gene environ-

ment pointed to different roles for the MCOs depending on environmental conditions. *Ngr147* seemed to be associated with carbohydrate modification, whereas *hh01-482* appeared to have multiple purposes due to the variety of surrounding gene functions. All MCO genes except for *ngr147* were linked to copper binding and/or transport.

Genes in proximity to each other are usually expressed consecutively with a likelihood of belonging to the same process. Close to *hh01-218* and *hh01-482* as well as *ngr578* and *ngr688*, there were genes located with membrane binding and transport function, indicating a large potential to bind copper ions and remove the resulting metalloprotein complex from the bacterial cell.

4.8 Protein expression and purification

By using auto-induction medium, the cell density of overnight cultures reached solid OD₆₀₀ values of 10-15. The toxicity of high copper concentrations was bypassed by ensuring a relatively high cell density prior to induction. This way, the addition of copper was delayed until absolutely necessary while afterwards it was rapidly integrated into the protein structures, decreasing the amount of free copper a lot faster than done by less dense cell cultures.

Interestingly, the copper-containing enzymes' blue shade was already slightly visible in pelletized expression cultures containing viable protein (Figure 32B). The striking color was even more distinctly obvious in the purification process, particularly when the protein gradually migrated through the column during elution. This represents an enormous advantage in comparison to uncolored proteins because the production of viable enzyme can be confirmed prior to purification, saving time and effort in case of expression issues.

A comparison of the experimental steps successfully executed with each putative MCO gene was visualized Table 40. Overall, it was accomplished to express eight of the fourteen identified putative MCOs in either *E. coli* BL21(DE3) or T7 SHuffle. Six of the gene products showed enzyme activity on ABTS and the four most promising candidates were further characterized. Given the fairly small number of initially identified putative MCO genes, this amounts to an excellent yield of viable enzymes – especially regarding metagenomic origin. Five of the six novel MCO genes from

lps typographus showed at least small activity on ABTS while three of them appeared suitable for enzyme characterization.

4.9 Enzyme characterization

4.9.1 Structural properties

Regarding protein structures, the classification of the four characterized MCOs by LccED showed that HH01-482 most likely contained two domains arranged in homotrimeric form while the metagenomic enzymes were all constructed from three domains potentially assembling to homodimers or staying in monomeric form. This was confirmed by using the SWISS-MODEL workspace to predict the enzymes' secondary and tertiary structures by using the template with the most identical AA sequence available. HH01-482 was modeled after a homotrimeric MCO of metagenomic origin that was only 30.5 % similar in sequence, hinting at a meager number of identified crystal structures of 2dMCOs.

In comparison to the template protein, the predicted structure of HH01-482 was a lot more condensed with less side chains or protruding elements. Together with the signal peptide cleavage site between AA26 and 27 and the gene for membrane binding and/or transport upstream of HH01-482, this attests to the secretion of the fully assembled enzyme. A larger protein yield with the expression strain T7 SHuffle in comparison to BL21(DE3) was consequential for HH01-482, as there possibly are numerous disulfide bonds involved in combining three monomers.

The most probable protein structures of *lps2204*, *lps14138* and *lps24328* were modeled after the template of CueO, further attesting to their close relation as discussed in 4.7.1. Comparing the images displayed in QMEAN color scheme with the template protein, the loop directly above the methionine-rich helix stands out as the most noticeable difference between the structures. *lps14138* was phylogenetically closest related to CueO and supported this fact by displaying the smallest loop structure of the three metagenomic enzymes.

As described in the introduction section, prior research stated that the methionine-rich helix functions as a lid structure regulating access to the MNC center and therefore the enzyme's substrate binding potential (Roberts *et al.* 2002). There was also proof that CueO is able to turn Cu^+ to Cu^{2+} by oxidation at additional copper binding

sites on the helix itself. The position of the helix was described to be modified by a previously undetected regulatory loop structure lying loosely above it after its conformation was induced by access copper ions (Singh *et al.* 2011; Wang *et al.* 2018). It is very likely, that the loop structures displayed by the metagenomic enzymes serve the same modulating purpose and help regulate the access to the enzymes' substrate binding sites. As the loops of the three enzymes each display a different pattern, it can be assumed that their native substrates all slightly differ, increasing the bark beetle's metabolic potential towards diverse phenolic compounds.

4.9.2 Storage stability

All four enzymes displayed above 80 % of residual activity on ABTS after 28 days of storage at -20 °C and 50 % glycerol with Ips14138 even showing a distinct increase to about 115 % after 7 days of incubation falling to a final value of 104 % at the end. Previous research assessed three MCO from *Bacillus* origin at 60-80 % of residual activity after 13 days under these storage conditions as highly stable (Ihsen *et al.* 2015). Two of the enzymes from the *Ips typographus* metagenome – Ips2204 and Ips14138 – had 97 % and 107 % left after 14 days, respectively. Following 28 days of storage at -20 °C and 50 % glycerol, Ips2204 preserved 92 % of initial activity while Ips14138 still performed better than at the start of the experiment with 104 %, rating significantly better than the before mentioned *Bacillus* MCOs.

Regarding storage at 4 °C, an MCO from *Sinorhizobium meliloti* retained >92 % of its activity after 7 days of storage (Pawlik *et al.* 2016) while Ips2204 and Ips14138 still showed 88 % and 94 %, respectively, performing at an equal level. Ips24328 and HH01-482 lost their activity more rapidly at the same temperature. The half-life of the *S. meliloti* MCO at 4 °C was reported to be reached after 21 days of incubation while Ips2204 and Ips14138 still had 78 % and 88 %, each. While Ips2204 still had 40 % of initial activity after 14 days at 22 °C, comparable to the *Bacillus* MCO (40-50 % after 25 °C for 13 days), the other tree *Ips* MCO performed worse with <30 % remaining.

These results showed that the activity of Ips2204 and Ips14138 exceeded several other bacterial enzymes at storage conditions of -20 °C and 50 % glycerol while they equal activity rates at 4 °C with a far slower loss of activity over time. However, only Ips2204 kept a decent activity of about 40 % after incubation at room temperature.

Ips14138 decreased to below 30 % while Ips24328 and HH01-482, both dropped to <10 %.

Overall, two of the newfound metagenomic MCOs performed considerably well regarding storage stability at different temperature and glycerol conditions, especially compared to other bacterial enzymes, representing a large benefit for industrial application.

4.9.3 Optimum pH

The metagenomic MCOs displayed an optimum for ABTS between 3.6 pH for Ips24328 and 4.4 pH for Ips2204 and Ips14138 while the strongest activity for HH01-482 was measured at 5.6 pH. Compared to literature, these results were not uncommon for bacterial MCOs, as they exhibit a broad variety of pH optima on ABTS ranging between pH 2.5 and 7 (Chauhan *et al.* 2017). The pH optimum for fungal MCOs on ABTS typically remains more in the acidic range between 2.0 and 5.5 pH (Baldrian 2006), being more restricted than their bacterial counterparts.

However, the range of pH over which Ips14138 still showed an activity above 40 % was comparably wide with 2.8 pH points. This points to a variety of possible industrial applications where it would still perform with a fairly high conversion rate.

4.9.4 Optimum temperature

Investigations of the MCOs temperature optimum were carried out on ABTS between 15 and 75 °C. Results showed the highest activity for Ips24328 at 25 °C and for HH01-482 at 27 °C. Ips2204 and Ips14138 were most active at 45 °C and 60 °C, respectively. The optimal temperature for fungal and most other laccase-like MCOs is known to lie between 50 and 70 °C (Baldrian 2006; Christopher *et al.* 2014) while there were bacterial laccases discovered with their highest activity on ABTS at 25 or up to 90 °C, the latter originating from most thermophilic bacteria like *Thermus thermophilus* (Chauhan *et al.* 2017). In comparison, Ips24328 and HH01-482 correlated with the lowest end of the known range, being most suitable for applications in the mesophilic scope. The maxima of Ips2204 (45 °C) and Ips14138 (60 °C) lay more in the expected span of optimum temperatures for bacterial MCOs.

All four enzymes exhibited above 90 % of their maximum activity over a range of 13-14 °C with their slopes declining slowly, indicating a wide variety of temperatures they were still amply operational at.

4.9.5 Heat stability

Even though its optimum reaction temperature was measured at 45 °C, Ips2204 appeared to be the least heat stable MCO tested. Its enzyme activity fell below 10 % after 30 s at 95 °C, 120 s at 80 °C and 120 min at 65 °C. Ips 24328 only performed slightly better at 80 °C by declining more slowly but otherwise presenting nearly the same pattern as Ips2204 at 95 and 65 °C.

The second most resistant enzyme turned out to be HH01-482. It showed the same activity after 30 s at 95 °C as without thermal treatment, before falling below 20 % after 30 more seconds. At 80 and 65 °C, HH01-482 even rose to approximately 130 and 125 %, respectively, and then declined similar to Ips2204 and Ips24328. Ips14138 not only demonstrated the highest optimum reaction temperature with 60 °C, but it additionally turned out to be the most heat stable MCO identified over the course of this study. It also displayed a rise in activity after thermal treatment even higher than HH01-482 and following incubation at all three temperatures.

Like HH01-482, it still exhibited 20 % of its non-heat-treated activity after 60 s of incubation at 95 °C. At the lower two temperatures however, Ips14138 performed far better than the other enzymes. After 300 s at 80 °C, it still displayed the same activity as without pre-incubation while its half-life was reached following about 420 s under these conditions. At 65 °C, Ips14138 still had 100 % of activity after 100 min of impact while it dropped to 50 % with approximately 170 min of incubation.

Previous research indicated a generally higher thermal stability for bacterial MCOs compared to their counterparts from fungal origin (Christopher *et al.* 2014). In comparison to CueO and other MCOs from γ -Proteobacteria, the heat stability of HH01-482 ranged on a similar level while Ips14138 performed significantly better. It withstood high temperatures to a degree comparable to several MCOs from *Bacillus* origin, which are known to frequently exhibit high heat resistance (Zeng *et al.* 2011; Ihssen *et al.* 2015; Chauhan *et al.* 2017).

Regarding thermal activation of Ips14138, there was an increase in activity to 160 % at 95 °C, to 246 % at 80 °C and to 250 % at 65 °C, signifying a boost up to 2.5 times the level of the untreated enzyme. In literature, there are also examples of thermal activation in CotA-like MCOs known from *Bacillus clausii* by factor 2 (Ihssen *et al.* 2015) and from *Bacillus spec.* by factor 4 (Mohammadian *et al.* 2010). As CotA – the most prominent *Bacillus* MCO – exhibits a high resistance against heat due to its function in spore coat formation (Enguita *et al.* 2003), the 2.5 fold heat activation of Ips14138 appears to be rather high for an enterobacterial enzyme. Thermal activation in MCOs is not yet completely understood but presumably attributed to the removal of inhibiting molecules from substrate binding regions or to conformational changes of the protein due to rising temperature (Brander *et al.* 2014).

This characteristic of both Ips14138 and HH01-482 can potentially be used in future biochemical applications by short-term boosting their activity to highly elevated levels.

4.10 Conclusion

In summary, the gut of the wood-feeding beetle *Ips typographus* was an excellent source for novel metagenomic enzymes, as previously expected. Probably due to the insects specialized lifestyle and nutrition, there was only a limited number of different microbial species uncovered and the identified laccase-like MCOs were all enterobacterial in origin. Regardless of the high resemblance in protein structure, there were still huge differences in storage, pH and temperature properties of the three characterized enzymes. This shows, that even small deviations like the loop pattern above the methionine-rich helix, can lead to completely different enzyme characteristics.

All four characterized enzymes could be expressed and purified with remarkably good amounts of protein while they were storable at -20 °C with 50 % glycerol over one month without losing more than 20 % of their initial activity. Ips2204 and Ips14138 were also performing comparably well at -20 °C with 25% glycerol, at 4 °C and Ips2204 even at 22 °C. Although, there were no distinct anomalies in pH optima, both Ips24328 and HH01-482 were most active at temperatures comparably low for bacterial and fungal MCOs (Baldrian 2006; Christopher *et al.* 2014). This could represent an advantage for application in biochemical reactions only operating at colder

temperatures. Also, all four enzymes displayed ample activity over a broad temperature range, indicating the ability to be utilized under a variety of different conditions. Initially, *lps14138* was thought to be less promising than other candidates due to its close phylogenetic relation to CueO from *E. coli*. Surprisingly, after further investigation it exhibited a superior heat stability and a strong thermal activation effect unknown for CueO but commonly identified in MCOs from *Bacillus* origin (Zeng *et al.* 2011; Chauhan *et al.* 2017). Combined with a high protein yield during expression and an excellent storage stability, *lps14138* fulfilled essential requirements for industrial utilization making it an excellent candidate for continuing research.

4.11 Outlook

To further strengthen the results from this study, additional enzyme characterization experiments should be performed in the future. Criteria like activity on different substrates, the ability to modify lignin directly or in the presence of laccase mediators as well as enzyme kinetics would be worthwhile research topics.

Transcriptome studies of the *lps typographus* intestines would also be extremely valuable to reveal which of the identified MCOs really were expressed in the beetle. Moreover, the assembled metagenomic dataset can be used to mine for vast numbers of proteins with a variety of different functions in the future. Especially enzymes associated with carbohydrate and other kinds of metabolism were shown to be elevated in the bark beetle compared to metagenomic datasets of different intestinal origins. There is still an enormous metabolic potential waiting to be discovered in the *lps typographus* metagenome.

References

- Abad ZG, Abad JA, Coffey MD, Oudemans PV, Man in 't Veld WA, Gruyter H de, Cunnington J, Louws FJ (2008): *Phytophthora bisheria* sp. nov., a new species identified in isolates from the Rosaceous raspberry, rose and strawberry in three continents. *Mycologia* **100**(1):99–110. doi:10.3852/mycologia.100.1.99
- Akter M, Inoue C, Komori H, Matsuda N, Sakurai T, Kataoka K, Higuchi Y, Shibata N (2016): Biochemical, spectroscopic and X-ray structural analysis of deuterated multicopper oxidase CueO prepared from a new expression construct for neutron crystallography. *Acta crystallographica. Section F, Structural biology communications* **72**(Pt 10):788–794. doi:10.1107/S2053230X1601400X
- Aneberg J, Karlsson CMG, Divne A-M, Bergin C, Homa F, Lindh MV, Hugerth LW, Ettema TJG, Bertilsson S, Andersson AF, Pinhassi J (2018): Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome* **6**. doi:10.1186/s40168-018-0550-0
- Altmann I (2005): Fraßgänge - Buchdrucker *Ips typographus* (Photograph: Bayerischer Wald - Hoher Bogen). Available at: <http://www.golddistel.de/kaefer/scolytidae/ips-typographus.htm>. By courtesy of (c) Ingrid Altmann.
- Altmann I (2008): Imago - Buchdrucker *Ips typographus* (Photograph: Bayerischer Wald - Furth am Wald). Available at: <http://www.golddistel.de/kaefer/scolytidae/ips-typographus.htm>. By courtesy of (c) Ingrid Altmann.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. *Journal of Molecular Biology* **215**(3):403–410. doi:10.1016/S0022-2836(05)80360-2
- Amann RI, Ludwig W, Schleifer KH (1995): Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**(1):143–169.
- Asano T, Seto Y, Hashimoto K, Kurushima H (2019): Mini-review an insect-specific system for terrestrialization: Laccase-mediated cuticle formation. *Insect biochemistry and molecular biology* **108**:61–70. doi:10.1016/j.ibmb.2019.03.007
- Avison MB, Heldreich CJ von, Higgins CS, Bennett PM, Walsh TR (2000): A TEM-2beta-lactamase encoded on an active Tn1-like transposon in the genome of a clinical isolate of *Stenotrophomonas maltophilia*. *The Journal of antimicrobial chemotherapy* **46**(6):879–884.
- Baldrian P (2006): Fungal laccases - occurrence and properties. *FEMS microbiology reviews* **30**(2):215–242. doi:10.1111/j.1574-4976.2005.00010.x
- Benkert P, Biasini M, Schwede T (2011): Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics (Oxford, England)* **27**(3):343–350. doi:10.1093/bioinformatics/btq662
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank. *Nucleic acids research* **28**(1):235–242. doi:10.1093/nar/28.1.235

- Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T (2017): Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific reports* **7**(1):10480. doi:10.1038/s41598-017-09654-8
- Bertrand T, Jolival C, Briozzo P, Caminade E, Joly N, Madzak C, Mougin C (2002): Crystal structure of a four-copper laccase complexed with an arylamine: insights into substrate recognition and correlation with kinetics. *Biochemistry* **41**(23):7325–7333. doi:10.1021/bi0201318
- Bienert S, Waterhouse A, Beer TAP de, Tauriello G, Studer G, Bordoli L, Schwede T (2017): The SWISS-MODEL Repository-new features and functionality. *Nucleic acids research* **45**(D1):D313-D319. doi:10.1093/nar/gkw1132
- Bolger AM, Lohse M, Usadel B (2014): Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**(15):2114–2120. doi:10.1093/bioinformatics/btu170
- Bradford MM (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**:248–254.
- Brander S, Mikkelsen JD, Kepp KP (2014): Characterization of an alkali- and halide-resistant laccase expressed in *E. coli*: CotA from *Bacillus clausii*. *PLoS one* **9**(6):e99402. doi:10.1371/journal.pone.0099402
- Buchholz PCF, Vogel C, Reusch W, Pohl M, Rother D, Spieß AC, Pleiss J (2016): BioCatNet: A Database System for the Integration of Enzyme Sequences and Biocatalytic Experiments. *Chembiochem : a European journal of chemical biology* **17**(21):2093–2098. doi:10.1002/cbic.201600462
- Burke GR, Moran NA (2011): Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome biology and evolution* **3**:195–208. doi:10.1093/gbe/evr002
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009): BLAST+: architecture and applications. *BMC bioinformatics* **10**:421. doi:10.1186/1471-2105-10-421
- Cañas AI, Camarero S (2010): Laccases and their natural mediators: biotechnological tools for sustainable eco-friendly processes. *Biotechnology advances* **28**(6):694–705. doi:10.1016/j.biotechadv.2010.05.002
- Ceja-Navarro JA, Karaoz U, Bill M, Hao Z, White RA, Arellano A, Ramanculova L, Filley TR, Berry TD, Conrad ME, Blackwell M, Nicora CD, Kim Y-M, Reardon PN, Lipton MS, Adkins JN, Pett-Ridge J, Brodie EL (2019): Gut anatomical properties and microbial functional assembly promote lignocellulose deconstruction and colony subsistence of a wood-feeding beetle. *Nature microbiology* **4**(5):864–875. doi:10.1038/s41564-019-0384-y
- Chauhan PS, Goradia B, Saxena A (2017): Bacterial laccase: recent update on production, properties and industrial applications. *3 Biotech* **7**(5):323. doi:10.1007/s13205-017-0955-7
- Christiansen E, Bakke A (1988): The Spruce Bark Beetle of Eurasia. In 'Dynamics of forest insect populations: Patterns, causes, implications'. (Ed. AA Berryman) pp. 479–503. (Springer-Verlag New York: Boston, MA)
- Christopher LP, Yao B, Ji Y (2014): Lignin Biodegradation with Laccase-Mediator Systems. *Frontiers in Energy Research* **2**:169. doi:10.3389/fenrg.2014.00012

- Corsaro D, Wylezich C, Venditti D, Michel R, Walochnik J, Wegensteiner R (2019): Filling gaps in the microsporidian tree: rDNA phylogeny of Chytridiopsis tytopographi (Microsporidia: Chytridiopsida). *Parasitology research* **118**(1):169–180. doi:10.1007/s00436-018-6130-1
- Crossman LC, Gould VC, Dow JM, Vernikos GS, Okazaki A, Sebahia M, Saunders D, Arrowsmith C, Carver T, Peters N, Adlem E, Kerhornou A, Lord A, Murphy L, Seeger K, Squares R, Rutter S, Quail MA, Rajandream M-A, Harris D, Churcher C, Bentley SD, Parkhill J, Thomson NR, Avison MB (2008): The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome biology* **9**(4):R74. doi:10.1186/gb-2008-9-4-r74
- Delalibera I, Vasanthakumar A, Burwitz B, Schloss P, Klepzig K, Handelsman J, Raffa K (2007): Composition of the bacterial community in the gut of the pine engraver, *Ips pini* (Say) (Coleoptera) colonizing red pine. *Symbiosis* **43**:97–104.
- Dubé E, Shareck F, Hurtubise Y, Daneault C, Beauregard M (2008): Homologous cloning, expression, and characterisation of a laccase from *Streptomyces coelicolor* and enzymatic decolourisation of an indigo dye. *Applied microbiology and biotechnology* **79**(4):597–603. doi:10.1007/s00253-008-1475-5
- Durand A-A, Bergeron A, Constant P, Buffet J-P, Déziel E, Guertin C (2015): Surveying the endomicrobiome and ectomicrobiome of bark beetles: The case of *Dendroctonus simplex*. *Scientific reports* **5**:17190. doi:10.1038/srep17190
- Eggert C, Temp U, Dean JFD, Eriksson K-EL (1996): A fungal metabolite mediates degradation of non-phenolic lignin structures and synthetic lignin by laccase. *FEBS letters* **391**(1-2):144–148. doi:10.1016/0014-5793(96)00719-3
- Eichlerová I, Šnajdr J, Baldrian P (2012): Laccase activity in soils: considerations for the measurement of enzyme activity. *Chemosphere* **88**(10):1154–1160. doi:10.1016/j.chemosphere.2012.03.019
- Engel P, Moran NA (2013): The gut microbiota of insects - diversity in structure and function. *FEMS microbiology reviews* **37**(5):699–735. doi:10.1111/1574-6976.12025
- Enguita FJ, Martins LO, Henriques AO, Carrondo MA (2003): Crystal structure of a bacterial endospore coat component. A laccase with enhanced thermostability properties. *The Journal of biological chemistry* **278**(21):19416–19425. doi:10.1074/jbc.M301251200
- Fang W, Fang Z, Zhou P, Chang F, Hong Y, Zhang X, Peng H, Xiao Y (2012): Evidence for lignin oxidation by the giant panda fecal microbiome. *PloS one* **7**(11):e50312. doi:10.1371/journal.pone.0050312
- Flechtmann CA, Ottati AL, Berisford CW (2000): Comparison of four trap types for ambrosia beetles (Coleoptera, Scolytidae) in Brazilian Eucalyptus stands. *Journal of economic entomology* **93**(6):1701–1707.
- Geib SM, Filley TR, Hatcher PG, Hoover K, Carlson JE, Jimenez-Gasco MdM, Nakagawa-Izumi A, Sleighter RL, Tien M (2008): Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences of the United States of America* **105**(35):12932–12937. doi:10.1073/pnas.0805257105

- Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G (2010): Laccases: a never-ending story. *Cellular and molecular life sciences : CMLS* **67**(3):369–385. doi:10.1007/s00018-009-0169-1
- Guest TC, Rashid S (2016): Anticancer Laccases: A Review. *Journal of Clinical & Experimental Oncology* **05**(01). doi:10.4172/2324-9110.1000153
- Guex N, Peitsch MC, Schwede T (2009): Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**:S162-73. doi:10.1002/elps.200900140
- Güllert S, Fischer MA, Turaev D, Noebauer B, Ilmberger N, Wemheuer B, Alawi M, Rattei T, Daniel R, Schmitz RA, Grundhoff A, Streit WR (2016): Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnology for biofuels* **9**:121. doi:10.1186/s13068-016-0534-x
- Haack FS, Poehlein A, Kröger C, Voigt CA, Piepenbring M, Bode HB, Daniel R, Schäfer W, Streit WR (2016): Molecular Keys to the *Janthinobacterium* and *Duganella* spp. Interaction with the Plant Pathogen *Fusarium graminearum*. *Frontiers in microbiology* **7**:1668. doi:10.3389/fmicb.2016.01668
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998): Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* **5**(10):R245-R249. doi:10.1016/S1074-5521(98)90108-9
- He F, Machemer-Noonan K, Golfier P, Unda F, Dechert J, Zhang W, Hoffmann N, Samuels L, Mansfield SD, Rausch T, Wolf S (2019): The in vivo impact of MsLAC1, a Miscanthus laccase isoform, on lignification and lignin composition contrasts with its in vitro substrate preference. *BMC plant biology* **19**(1):552. doi:10.1186/s12870-019-2174-3
- Hiraoka S, Yang C-c, Iwasaki W (2016): Metagenomics and Bioinformatics in Microbial Ecology: Current Status and Beyond. *Microbes and Environments* **31**(3):204–212. doi:10.1264/jsme2.ME16024
- Hoegger PJ, Kilaru S, James TY, Thacker JR, Kües U (2006): Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *The FEBS journal* **273**(10):2308–2326. doi:10.1111/j.1742-4658.2006.05247.x
- Holterman M, Karegar A, Mooijman P, van Megen H, van den Elsen S, Vervoort MTW, Quist CW, Karssen G, Decraemer W, Opperman CH, Bird DM, Kam-menga J, Goverse A, Smant G, Helder J (2017): Disparate gain and loss of parasitic abilities among nematode lineages. *PloS one* **12**(9). doi:10.1371/journal.pone.0185445
- Hornung C, Poehlein A, Haack FS, Schmidt M, Dierking K, Pohlen A, Schulenburg H, Blokesch M, Plener L, Jung K, Bonge A, Krohn-Molt I, Utpatel C, Timmermann G, Spieck E, Pommerening-Röser A, Bode E, Bode HB, Daniel R, Schmeisser C, Streit WR (2013): The *Janthinobacterium* sp. HH01 genome encodes a homologue of the *V. cholerae* CqsA and *L. pneumophila* LqsA autoinducer synthases. *PloS one* **8**(2):e55045. doi:10.1371/journal.pone.0055045

- Hu Q, Min L, Yang X, Jin S, Zhang L, Li Y, Ma Y, Qi X, Li D, Liu H, Lindsey K, Zhu L, Zhang X (2018): Laccase GhLac1 Modulates Broad-Spectrum Biotic Stress Tolerance via Manipulating Phenylpropanoid Pathway and Jasmonic Acid Synthesis. *Plant physiology* **176**(2):1808–1823. doi:10.1104/pp.17.01628
- Hublik G, Schinner F (2000): Characterization and immobilization of the laccase from *Pleurotus ostreatus* and its use for the continuous elimination of phenolic pollutants [Online]. *Enzyme and Microbial Technology* **27**(3-5):330–336. doi:10.1016/S0141-0229(00)00220-9 Available at: <https://www.sciencedirect.com/science/article/pii/S0141022900002209?via%3Dihub> (verified 30 May 2019)
- Hui F-L, Chen L, Chu X-Y, Niu Q-H, Ke T (2013): *Wickerhamomyces mori* sp. nov., an anamorphic yeast species found in the guts of wood-boring insect larvae. *International journal of systematic and evolutionary microbiology* **63**(Pt 3):1174–1178. doi:10.1099/ijs.0.048637-0
- Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC (2015): The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in genomic sciences* **10**:86. doi:10.1186/s40793-015-0077-y
- Ihsen J, Reiss R, Luchsinger R, Thöny-Meyer L, Richter M (2015): Biochemical properties and yields of diverse bacterial laccase-like multicopper oxidases expressed in *Escherichia coli*. *Scientific reports* **5**:10465. doi:10.1038/srep10465
- Ilmberger N, Güllert S, Dannenberg J, Rabausch U, Torres J, Wemheuer B, Alawi M, Poehlein A, Chow J, Turaev D, Rattei T, Schmeisser C, Salomon J, Olsen PB, Daniel R, Grundhoff A, Borchert MS, Streit WR (2014): A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. *PloS one* **9**(9):e106707. doi:10.1371/journal.pone.0106707
- John R (2016): Fallenstellerinnen in Pheromonwolken? – Borkenkäfermonitoring ist immer en vogue! *FVA-einblick* **1**:10–14.
- Jones RT, Sanchez LG, Fierer N (2013): A cross-taxon analysis of insect-associated bacterial diversity. *PloS one* **8**(4):e61218. doi:10.1371/journal.pone.0061218
- Kanehisa M (1997): Linking databases and organisms: GenomeNet resources in Japan. *Trends in biochemical sciences* **22**(11):442–444.
- Kanehisa M, Goto S, Kawashima S, Nakaya A (2002): The KEGG databases at GenomeNet. *Nucleic acids research* **30**(1):42–46. doi:10.1093/nar/30.1.42
- Kelsey RG, Beh MM, Shaw DC, Manter DK (2013): Ethanol attracts scolytid beetles to *Phytophthora ramorum* cankers on coast live oak. *Journal of chemical ecology* **39**(4):494–506. doi:10.1007/s10886-013-0271-6
- Kim Y, Maltseva N, Shatsman S, Anderson WF, Joachimiak A (2012): Crystal Structure of Uncharacterized Cupredoxin-like Domain Protein Cupredoxin_1 with Copper Bound from *Bacillus anthracis*: Protein Data Bank ID: 4HCF.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013): Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research* **41**(1):e1. doi:10.1093/nar/gks808

- Komori H, Miyazaki K, Higuchi Y (2009): X-ray structure of a two-domain type laccase: a missing link in the evolution of multi-copper proteins. *FEBS letters* **583**(7):1189–1195. doi:10.1016/j.febslet.2009.03.008
- Köpcke B, Weber RWS, Anke H (2002): Galiellalactone and its biogenetic precursors as chemotaxonomic markers of the Sarcosomataceae (Ascomycota). *Phytochemistry* **60**(7):709–714. doi:10.1016/S0031-9422(02)00193-0
- Krishnan M, Bharathiraja C, Pandiarajan J, Prasanna VA, Rajendhran J, Gunasekaran P (2014): Insect gut microbiome - An unexploited reserve for biotechnological application. *Asian Pacific journal of tropical biomedicine* **4**(Suppl 1):S16-21. doi:10.12980/APJTB.4.2014C95
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018): MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* **35**(6):1547–1549. doi:10.1093/molbev/msy096
- Kyrpides NC (1999): Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. *Bioinformatics (Oxford, England)* **15**(9):773–774.
- Li C, Li D, Zhou H, Li J, Lu S (2019): Analysis of the laccase gene family and miR397-/miR408-mediated posttranscriptional regulation in *Salvia miltiorrhiza*. *PeerJ* **7**:e7605. doi:10.7717/peerj.7605
- Li W, Godzik A (2006): Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* **22**(13):1658–1659. doi:10.1093/bioinformatics/btl158
- Lobinger G, Krüger F, Stahl F (2016): Buchdrucker und Kupferstecher an Fichte: Bayerische Landesanstalt für Wald und Forstwirtschaft (LWF): Merkblatt) **14**.
- Lombardero MJ, Ayres MP, Krivak-Tetley FE, Fitz KNE (2016): Population biology of the European woodwasp, *Sirex noctilio*, in Galicia, Spain. *Bulletin of entomological research* **106**(5):569–580. doi:10.1017/S0007485316000043
- Machczynski MC, Vijgenboom E, Samyn B, Canters GW (2004): Characterization of SLAC: a small laccase from *Streptomyces coelicolor* with unprecedented activity. *Protein science : a publication of the Protein Society* **13**(9):2388–2397. doi:10.1110/ps.04759104
- Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012): IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic acids research* **40**:D115-22. doi:10.1093/nar/gkr1044
- Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC (2014): IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic acids research* **42**(Database issue):D560-7. doi:10.1093/nar/gkt963
- Maseko B, Burgess TI, Coutinho TA, Wingfield MJ (2007): Two new *Phytophthora* species from South African Eucalyptus plantations. *Mycological research* **111**(Pt 11):1321–1338. doi:10.1016/j.mycres.2007.08.011

- Mason CJ, Scully ED, Geib SM, Hoover K (2016): Contrasting diets reveal metabolic plasticity in the tree-killing beetle, *Anoplophora glabripennis* (Cerambycidae: Lamiinae). *Scientific reports* **6**:33813. doi:10.1038/srep33813
- Mate DM, Alcalde M (2015): Laccase engineering: from rational design to directed evolution. *Biotechnology advances* **33**(1):25–40. doi:10.1016/j.biotechadv.2014.12.007
- McCutcheon JP, Moran NA (2011): Extreme genome reduction in symbiotic bacteria. *Nature reviews. Microbiology* **10**(1):13–26. doi:10.1038/nrmicro2670
- Messerschmidt A (1997): Multi-copper oxidases. (World Scientific Pub. Co: Singapore, River Edge, N.J)
- Messerschmidt A (2010): Copper Metalloenzymes. In 'Comprehensive natural products II: Chemistry and biology'. (Ed. L Mander) pp. 489–545. (Elsevier: Amsterdam)
- Messerschmidt A, Ladenstein R, Huber R, Bolognesi M, Avigliano L, Petruzzelli R, Rossi A, Finazzi-Agró A (1992): Refined crystal structure of ascorbate oxidase at 1.9 Å resolution. *Journal of Molecular Biology* **224**(1):179–205. doi:10.1016/0022-2836(92)90583-6
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008): The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* **9**:386. doi:10.1186/1471-2105-9-386
- Mohammadian M, Fathi-Roudsari M, Mollania N, Badoei-Dalfard A, Khajeh K (2010): Enhanced expression of a recombinant bacterial laccase at low temperature and microaerobic conditions: purification and biochemical characterization. *Journal of industrial microbiology & biotechnology* **37**(8):863–869. doi:10.1007/s10295-010-0734-5
- Montfort WR, Roberts SA, Singh SK (2011): The reduced form of CueO: Protein Data Bank ID: 3PAV.
- Morozova OV, Shumakovich GP, Gorbacheva MA, Shleev SV, Yaropolov AI (2007): "Blue" laccases. *Biochemistry (Moscow)* **72**(10):1136–1150. doi:10.1134/S0006297907100112
- Mot AC, Silaghi-Dumitrescu R (2012): Laccases: complex architectures for one-electron oxidations. *Biochemistry. Biokhimiia* **77**(12):1395–1407. doi:10.1134/S0006297912120085
- Nakamura K, Kawabata T, Yura K, Go N (2003): Novel types of two-domain multi-copper oxidases: possible missing links in the evolution. *FEBS letters* **553**(3):239–244. doi:10.1016/S0014-5793(03)01000-7
- Niemeyer H, Schröder T, Watzek G (1983): Eine neue Lockstoff-Falle zur Bekämpfung von rinden- und holzbrütenden Borkenkäfern. *Der Forst- und Holzwirt* **38**:105–112.
- Notredame C, Higgins DG, Heringa J (2000): T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**(1):205–217. doi:10.1006/jmbi.2000.4042

- NW-FVA (Ed.) (2009): Integrierte Bekämpfung rindenbrütender Borkenkäfer: Praxis-Information Nr 1 - Februar 2009. (Nordwestdeutsche Forstliche Versuchsanstalt - Abteilung Waldschutz: Göttingen)
- Orzheshkovskiy VV, Trishchynska MA (2019): Ceruloplasmin: Its Role in the Physiological and Pathological Processes. *Neurophysiology* **51**(2):141–149. doi:10.1007/s11062-019-09805-9
- Pardo I, Rodríguez-Escribano D, Aza P, Salas F de, Martínez AT, Camarero S (2018): A highly stable laccase obtained by swapping the second cupredoxin domain. *Scientific reports* **8**(1):15669. doi:10.1038/s41598-018-34008-3
- Pawlik A, Wójcik M, Rułka K, Motyl-Gorzel K, Osińska-Jaroszuk M, Wielbo J, Marek-Kozaczuk M, Skorupska A, Rogalski J, Janusz G (2016): Purification and characterization of laccase from *Sinorhizobium meliloti* and analysis of the lacc gene. *International journal of biological macromolecules* **92**:138–147. doi:10.1016/j.ijbiomac.2016.07.012
- Peng Y, Leung HCM, Yiu SM, Chin FYL (2012): IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics (Oxford, England)* **28**(11):1420–1428. doi:10.1093/bioinformatics/bts174
- Peng Z, Dittmer NT, Lang M, Brummett LM, Braun CL, Davis LC, Kanost MR, Gorman MJ (2015): Multicopper oxidase-1 orthologs from diverse insect species have ascorbate oxidase activity. *Insect biochemistry and molecular biology* **59**:58–71. doi:10.1016/j.ibmb.2015.02.005
- Petersen TN, Brunak S, Heijne G von, Nielsen H (2011): SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**(10):785–786. doi:10.1038/nmeth.1701
- Piontek K, Antorini M, Choinowski T (2002): Crystal structure of a laccase from the fungus *Trametes versicolor* at 1.90-Å resolution containing a full complement of coppers. *The Journal of biological chemistry* **277**(40):37663–37669. doi:10.1074/jbc.M204571200
- Pourcel L, Routaboul J-M, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005): TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. *The Plant cell* **17**(11):2966–2980. doi:10.1105/tpc.105.035154
- Pruesse E, Peplies J, Glöckner FO (2012): SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics (Oxford, England)* **28**(14):1823–1829. doi:10.1093/bioinformatics/bts252
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013): The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**(Database issue):D590-6. doi:10.1093/nar/gks1219
- Reiss R, Ihssen J, Richter M, Eichhorn E, Schilling B, Thöny-Meyer L (2013): Laccase versus laccase-like multi-copper oxidase: a comparative study of similar enzymes with diverse substrate spectra. *PloS one* **8**(6):e65633. doi:10.1371/journal.pone.0065633
- Ricci I, Damiani C, Scuppa P, Mosca M, Crotti E, Rossi P, Rizzi A, Capone A, Gonella E, Ballarini P, Chouaia B, Sagnon N'f, Esposito F, Alma A, Mandrioli M,

- Sacchi L, Bandi C, Daffonchio D, Favia G (2011): The yeast *Wickerhamomyces anomalus* (*Pichia anomala*) inhabits the midgut and reproductive system of the Asian malaria vector *Anopheles stephensi*. *Environmental microbiology* **13**(4):911–921. doi:10.1111/j.1462-2920.2010.02395.x
- Roberts SA, Weichsel A, Grass G, Thakali K, Hazzard JT, Tollin G, Rensing C, Montfort WR (2002): Crystal structure and electron transfer kinetics of CueO, a multicopper oxidase required for copper homeostasis in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **99**(5):2766–2771. doi:10.1073/pnas.052710499
- Roger M, Biaso F, Castelle CJ, Bauzan M, Chaspoul F, Lojou E, Sciara G, Caffarri S, Giudici-Ortoni M-T, Ilbert M (2014): Spectroscopic Characterization of a Green Copper Site in a Single-Domain Cupredoxin. *PloS one* **9**(6). doi:10.1371/journal.pone.0098941
- Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW (2018): NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics (Oxford, England)* **34**(21):3755–3758. doi:10.1093/bioinformatics/bty419
- Sambrook J, Russell DW (2001): *Molecular cloning: A laboratory manual* (3rd edn). (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y)
- Samyginina VR, Sokolov AV, Bourenkov G, Petoukhov MV, Pulina MO, Zakharova ET, Vasilyev VB, Bartunik H, Svergun DI (2013): Ceruloplasmin: macromolecular assemblies with iron-containing acute phase proteins. *PloS one* **8**(7):e67145. doi:10.1371/journal.pone.0067145
- Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quéré A, Wollherr A, Heine-meyer I, Morgenstern B, Pommerening-Röser A, Flores M, Palacios R, Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW, Streit WR (2009): *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Applied and environmental microbiology* **75**(12):4035–4045. doi:10.1128/AEM.00515-09
- Schneider KP, Gewessler U, Flock T, Heinzle A, Schenk V, Kaufmann F, Sigl E, Guebitz GM (2012): Signal enhancement in polysaccharide based sensors for infections by incorporation of chemically modified laccase. *New biotechnology* **29**(4):502–509. doi:10.1016/j.nbt.2012.03.005
- Schroeder LM (2001): Tree Mortality by the Bark Beetle *Ips typographus* (L.) in storm-disturbed stands. *Integrated Pest Management Reviews* **6**(3-4):169–175. doi:10.1023/A:1025771318285
- Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, Glavina del Rio T, Chovatia M, Herr JR, Carlson JE (2013): Metagenomic profiling reveals lignocel-lulose degrading system in a microbial community associated with a wood-feed-ing beetle. *PloS one* **8**(9):e73827. doi:10.1371/journal.pone.0073827
- Sharma P, Goel R, Capalash N (2007): Bacterial laccases. *World Journal of Micro-biology and Biotechnology* **23**(6):823–832. doi:10.1007/s11274-006-9305-3
- Singh SK, Roberts SA, McDevitt SF, Weichsel A, Wildner GF, Grass GB, Rensing C, Montfort WR (2011): Crystal structures of multicopper oxidase CueO bound to copper(I) and silver(I): functional role of a methionine-rich sequence. *The Jour-nal of biological chemistry* **286**(43):37849–37857. doi:10.1074/jbc.M111.293589

- Sirim D, Wagner F, Wang L, Schmid RD, Pleiss J (2011): The Laccase Engineering Database: a classification and analysis system for laccases and related multi-copper oxidases. *Database : the journal of biological databases and curation* **2011**:bar006. doi:10.1093/database/bar006
- Skálová T, Dohnálek J, Østergaard LH, Østergaard PR, Kolenko P, Dusková J, Stepánková A, Hasek J (2009): The structure of the small laccase from *Streptomyces coelicolor* reveals a link between laccases and nitrite reductases. *Journal of Molecular Biology* **385**(4):1165–1178. doi:10.1016/j.jmb.2008.11.024
- Sponheimer M, Clauss M, Codron D (2019): Dietary Evolution: The Panda Paradox. *Current biology : CB* **29**(11):R417-R419. doi:10.1016/j.cub.2019.04.045
- Stefanini I (2018): Yeast-insect associations: It takes guts. *Yeast (Chichester, England)* **35**(4):315–330. doi:10.1002/yea.3309
- Studier FW (2005): Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**(1):207–234. doi:10.1016/j.pep.2005.01.016
- Susoy V, Herrmann M (2014): Preferential host switching and codivergence shaped radiation of bark beetle symbionts, nematodes of *Micoletzkyia* (Nematoda: Diplogastridae). *Journal of evolutionary biology* **27**(5):889–898. doi:10.1111/jeb.12367
- Tepper AWJW, Aartsma TJ, Canters GW (2011): Channeling of electrons within SLAC, the small laccase from *Streptomyces coelicolor*. *Faraday Discuss* **148**(6885):161–171. doi:10.1039/C002585B
- Thurston CF (1994): The structure and function of fungal laccases. *Microbiology (Reading, England)* **140**(1):19–26. doi:10.1099/13500872-140-1-19
- Tonka T, Weiser J (2010): Budding: a new stage in the development of *Chytridiopsis typographi* (Zygomycetes: Microsporidia). *Journal of invertebrate pathology* **104**(1):17–22. doi:10.1016/j.jip.2010.01.003
- Trinick MJ (1980): Relationships amongst the fast-growing *Rhizobia* of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* spp, *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with other rhizobial groups. *Journal of Applied Bacteriology* **49**(1):39–53.
- Urbina H, Schuster J, Blackwell M (2013): The gut of Guatemalan passalid beetles: a habitat colonized by cellobiose- and xylose-fermenting yeasts. *Fungal Ecology* **6**(5):339–355. doi:10.1016/j.funeco.2013.06.005
- Vasanthakumar A, Delalibera I, Handelsman J, Klepzig KD, Schloss PD, Raffa KF (2006): Characterization of Gut-Associated Bacteria in Larvae and Adults of the Southern Pine Beetle, *Dendroctonus frontalis* Zimmermann. *Environmental Entomology* **35**(6):1710–1717. doi:10.1603/0046-225X(2006)35[1710:COG-BIL]2.0.CO;2
- Vollmers J, Wiegand S, Kaster A-K (2017): Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *PloS one* **12**(1). doi:10.1371/journal.pone.0169662
- Wallace IM, O'Sullivan O, Higgins DG, Notredame C (2006): M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic acids research* **34**(6):1692–1699. doi:10.1093/nar/gkl091

- Wang F, Xu L, Zhao L, Ding Z, Ma H, Terry N (2019): Fungal Laccase Production from Lignocellulosic Agricultural Wastes by Solid-State Fermentation: A Review. *Microorganisms* **7**(12). doi:10.3390/microorganisms7120665
- Wang H, Liu X, Zhao J, Yue Q, Yan Y, Gao Z, Dong Y, Zhang Z, Fan Y, Tian J, Wu N, Gong Y (2018): Crystal structures of multicopper oxidase CueO G304K mutant: structural basis of the increased laccase activity. *Scientific reports* **8**(1):14252. doi:10.1038/s41598-018-32446-7
- Wang L, Nie Y, Tang Y-Q, Song X-M, Cao K, Sun L-Z, Wang Z-J, Wu X-L (2016): Diverse Bacteria with Lignin Degrading Potentials Isolated from Two Ranks of Coal. *Frontiers in microbiology* **7**:1428. doi:10.3389/fmicb.2016.01428
- Wang Q, Li G, Zheng K, Zhu X, Ma J, Wang D, Tang K, Feng X, Leng J, Yu H, Yang S, Feng X (2019): The Soybean Laccase Gene Family: Evolution and Possible Roles in Plant Defense and Stem Strength Selection. *Genes* **10**(9). doi:10.3390/genes10090701
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, Beer TAP de, Rempfer C, Bordoli L, Lepore R, Schwede T (2018): SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**(W1):W296-W303. doi:10.1093/nar/gky427
- Wermelinger B (2004): Ecology and management of the spruce bark beetle *Ips typographus*—a review of recent research. *Forest Ecology and Management* **202**(1-3):67–82. doi:10.1016/j.foreco.2004.07.018
- White TC, Findley K, Dawson TL, Scheynius A, Boekhout T, Cuomo CA, Xu J, Saunders CW (2014): Fungi on the Skin: Dermatophytes and Malassezia. *Cold Spring Harbor Perspectives in Medicine* **4**(8). doi:10.1101/cshperspect.a019802
- Whitman WB, Coleman DC, Wiebe WJ (1998): Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* **95**(12):6578–6583.
- Witayakran S, Ragauskas AJ (2009): Synthetic Applications of Laccase in Green Chemistry. *Advanced Synthesis & Catalysis* **351**(9):1187–1209. doi:10.1002/adsc.200800775
- Yang X, Tyler BM, Hong C (2017): An expanded phylogeny for the genus *Phytophthora*. *IMA fungus* **8**(2):355–384. doi:10.5598/imafungus.2017.08.02.09
- Yin Q, Zhou G, Peng C, Zhang Y, Kües U, Liu J, Xiao Y, Fang Z (2019): The first fungal laccase with an alkaline pH optimum obtained by directed evolution and its application in indigo dye decolorization. *AMB Express* **9**(1):151. doi:10.1186/s13568-019-0878-2
- Yoshida H (1883): Chemistry of lacquer (urushi). Part I. Communication from the Chemical Society of Tokio. *Journal of the Chemical Society, Transactions*(43):472–486.
- Zeng J, Lin X, Zhang J, Li X, Wong MH (2011): Oxidation of polycyclic aromatic hydrocarbons by the bacterial laccase CueO from *E. coli*. *Applied microbiology and biotechnology* **89**(6):1841–1849. doi:10.1007/s00253-010-3009-1
- Zimbardi ALRL, Camargo PF, Carli S, Aquino Neto S, Meleiro LP, Rosa JC, Andrade AR de, Jorge JA, Furriel RPM (2016): A High Redox Potential Laccase from *Pycnoporus sanguineus* RP15: Potential Application for Dye Decolorization. *International journal of molecular sciences* **17**(5). doi:10.3390/ijms17050672

List of Abbreviations

2dMCO	two-domain multicopper oxidase
3dMCO	three-domain multicopper oxidase
AA	Amino acid
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
AO	Ascorbate oxidase
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CBM	Copper binding motif
CBS	Copper binding site
CopA	Copper efflux P-type ATPase
CotA	Spore coat protein A
CueO	Copper efflux oxidase
CV	Column volume
D	Domain
FPLC	Fast protein liquid chromatography
HF	Homologous family
IMG	Integrated Microbial Genomes database
LB	Luria Bertani medium
LCA	Lowest common ancestor algorithm
LccED	Laccase and Multicopper Oxidase Engineering Database
MCO	Multicopper oxidase
MCS	Multiple cloning site
MNC	Mononuclear copper center
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
ORF	Open reading frame
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SF	Superfamily
SLAC	Small laccase
T	Type
T _A	Annealing temperature
TEMED	Tetramethylethylenediamine
T _M	Melting temperature
TNC	Trinuclear copper center

Appendix

pET22b::StreptII

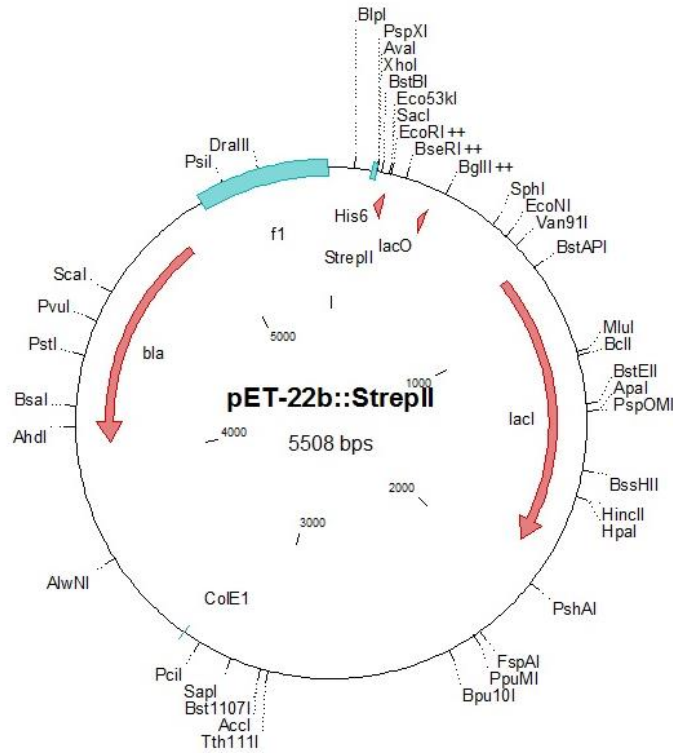


Figure 49: Vector map of pET22b::StreptII

pET22b-SDM::StreptII

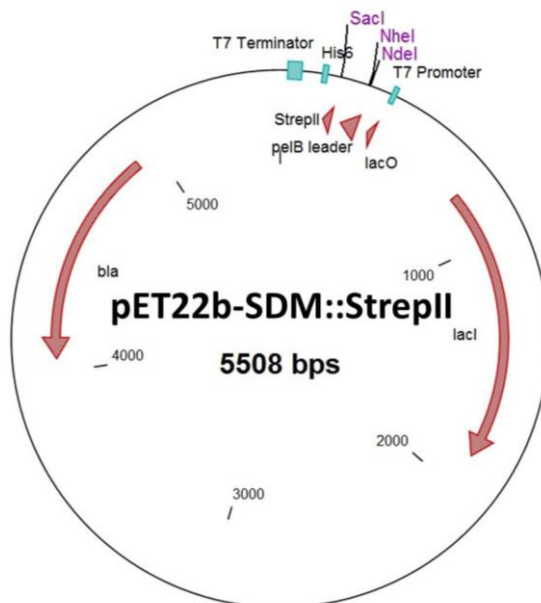


Figure 50: Vector map of pET22b-SDM::StreptII

Table 43: Raw data of *Ips typographus* 16S rRNA sequences

Class	Genus	Percentage [%]	Number of sequences
Actinobacteria	Propionibacterium	0,06	1
Bacilli	Bacillus	1,56	26
Bacilli	Planomicrobium	0,72	12
Bacilli	Geobacillus	0,12	2
Bacilli	Carnobacterium	0,06	1
α -Proteobacteria	Phyllobacterium	1,50	25
α -Proteobacteria	Ochrobactrum	0,12	2
α -Proteobacteria	Paracoccus	0,06	1
γ -Proteobacteria	Morganella	57,11	952
γ -Proteobacteria	Enterobacter	22,20	370
γ -Proteobacteria	Serratia	5,64	94
γ -Proteobacteria	Escherichia	3,72	62
γ -Proteobacteria	Hafnia	2,88	48
γ -Proteobacteria	Raoultella	0,78	13
γ -Proteobacteria	Providencia	0,72	12
γ -Proteobacteria	Pantoea	0,54	9
γ -Proteobacteria	Leminorella	0,48	8
γ -Proteobacteria	Yersinia	0,48	8
γ -Proteobacteria	B38	0,42	7
γ -Proteobacteria	Rahnella	0,30	5
γ -Proteobacteria	Cedecea	0,24	4
γ -Proteobacteria	Cronobacter	0,06	1
γ -Proteobacteria	Erwinia	0,06	1
γ -Proteobacteria	Acinetobacter	0,06	1
γ -Proteobacteria	Enhydrobacter	0,06	1
γ -Proteobacteria	Pseudomonas	0,06	1
Total		100,00	1667

Table 44: Phylogenetic distribution of protein-coding genes from several metagenomic datasets from the Integrated Microbial Genomes database (IMG) in percent.

	<i>I. typographus</i>	<i>A. glabripennis</i>	<i>V. sinuatoctollis</i>	<i>N. corniger</i>	<i>S. noctilio</i>	<i>A. mellifera</i>	<i>G. gallus</i>	<i>C. familiaris</i>	<i>A. melano-leuca</i>
Actinobacteria	12.73	10.89	7.61	16.36	7.08	9.39	12.74	8.77	2.90
Bacteroidetes	5.64	6.27	9.10	7.50	3.99	6.06	7.67	6.88	4.19
Cyanobacteria	2.40	2.73	1.76	2.37	2.28	2.41	2.65	2.34	3.24
Deinococcus-Thermus	0.55	3.25	0.45	0.58	0.73	0.57	0.77	0.56	0.11
Firmicutes	22.85	27.58	39.31	32.71	17.98	29.37	26.60	31.65	59.58
Fusobacteria	0.60	0.90	1.13	0.76	0.41	0.88	0.70	1.02	1.34
Proteobacteria	48.43	41.87	32.69	33.17	61.27	43.99	42.16	41.34	25.24
Spirochaetes	1.85	0.99	1.26	1.92	0.81	1.01	1.26	0.87	1.01
Tenericutes	1.05	0.66	1.40	1.08	0.41	1.76	1.51	1.48	0.67
Others	3.90	4.86	5.31	3.56	5.04	4.57	3.93	5.10	1.73

Table 45: Functional distribution of protein-coding genes from several metagenomic datasets according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) in percent.

	<i>I. tyograplus</i>	<i>A. glabripennis</i>	<i>V. sinuaticollis</i>	<i>N. corniger</i>	<i>S. noctilio</i>	<i>A. mellifera</i>	<i>G. gallus</i>	<i>C. familiaris</i>	<i>A. melanoleuca</i>
Amino acid metabolism	15.72	14.17	11.41	11.36	11.66	12.29	12.01	11.60	8.57
Biosynthesis of other secondary metabolites	1.66	1.64	0.98	1.68	1.23	1.35	1.75	1.28	0.90
Carbohydrate metabolism	17.24	14.59	9.83	14.72	10.96	15.36	15.20	14.44	12.69
Cell growth and death	0.75	1.04	2.17	1.42	1.20	1.20	1.42	1.23	1.64
Cell motility	0.27	0.86	0.76	1.07	2.70	0.95	0.60	0.41	2.10
Energy metabolism	9.10	8.12	4.51	6.86	6.38	7.33	8.37	9.13	8.00
Folding, sorting and degradation	2.34	2.55	4.51	4.23	2.85	2.92	3.24	3.94	3.02
Glycan biosynthesis and metabolism	2.59	1.80	2.51	2.42	1.72	2.94	3.71	4.45	2.31
Lipid metabolism	5.14	4.78	3.12	3.68	4.59	3.59	3.68	3.51	2.91
Membrane transport	2.54	8.06	5.97	6.76	6.77	10.03	7.31	4.76	9.22
Metabolism of cofactors and vitamins	8.86	6.35	3.98	5.12	5.65	7.33	6.60	7.25	5.82
Metabolism of other amino acids	4.88	4.21	2.06	3.68	3.73	3.68	3.27	2.75	2.89
Metabolism of terpenoids and polyketides	2.28	3.14	1.70	2.04	2.61	2.24	1.93	2.30	1.95
Nucleotide metabolism	7.25	5.99	4.66	7.78	4.68	7.16	7.65	8.64	10.25
Replication and repair	3.70	3.61	4.20	4.92	2.49	5.46	5.97	5.46	6.98
Signal transduction	3.61	3.67	5.29	3.92	7.61	3.63	3.40	3.03	3.83
Transcription	0.53	0.72	1.83	0.95	1.07	0.54	0.69	1.07	0.93
Translation	2.58	4.32	7.30	5.84	2.72	4.93	6.21	7.45	7.96
Transport and catabolism	1.05	0.82	2.72	1.13	1.52	0.55	0.65	1.05	0.28
Xenobiotics biodegradation and metabolism	3.73	5.01	1.45	2.42	4.14	2.90	2.05	2.20	1.70
Others	4.17	4.56	19.04	8.01	13.72	3.61	4.29	4.03	6.04

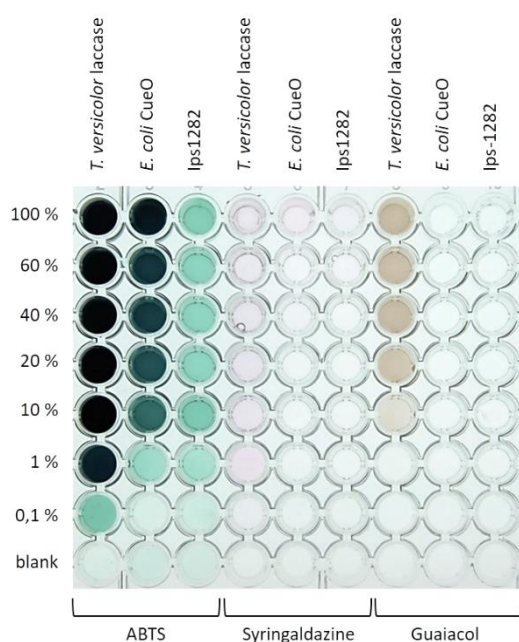


Figure 51: Comparative enzyme activity of three MCOs on ABTS, syringaldazine and guaiacol. This image demonstrated the activity of decreasing relative enzyme concentrations of *T. versicolor* laccase (TVL; 2 mg/ml), CueO from *E. coli* (0.6 mg/ml) and Ips1282 (crude cell extract) on ABTS, syringaldazine and guaiacol (cf. 2.7.6). The TVL showed the strongest activity on all three substrates even considering the higher protein concentration. CueO was highly active on ABTS, only weakly on syringaldazine and exhibited no detectable activity on guaiacol under these test conditions. The least substrate turnover was shown by Ips1282 which was only slightly active on ABTS and barely on syringaldazine. Despite differing protein concentrations, the ABTS activity assay appeared the most suitable for characterizing the kind of multicopper oxidases identified during this study.

CueO

Nucleotide sequence (1476 nt):

ATGGCAGAACGCCAACGTTACCGATCCCTGATTTGCTCACGACCGATGCCCCGTAATCGCATTCAGTTAACTA
TTGGCGCAGGCCAGTCCACCTTTGGCGGGAAAACGCAACTACCTGGGGCTATAACGGCAATCTGCTGGGGCC
GGCGGTGAAATTACAGCGCGCAAAGCGGTAACGGTTGATATCTACAACCAACTGACGGAAGAGACAACGTTG
CACTGGCACGGGCTGGAAGTACCGGGTGAAGTCGACGGCGGCCCGCAGGGAATTATTCGCCCAGGTGGCAAGC
GCTCGGTGACGTTGAACGTTGATCAACCTGCCGCTACCTGCTGGTTCCATCCGCATCAGCACGGCAAAACCGG
GCGACAGGTGGCGATGGGGCTGGCTGGGGCTGGTGGTGAATTGAAGATGACGAGATCCTGAAATTAATGCTGCCA
AAACAGTGGGGTATCGATGATGTTCCGGTGATCGTTTCCAGGATAAGAAATTTAGCGCCGACGGGCAGATTGATT
ATCAACTGGATGTGATGACCGCCGCCGTTGGGCTGGTTTGGCGATACGTTGCTGACCAACGGTGCAATCTACCC
GCAACACGCTGCCCCGCTGGTTGGCTGCGCCTGCGTTTGTCAATGGCTGTAATGCCCGTTTCGCTCAATTTT
GCCACCAGCGACAATCGCCCGCTGTATGTGATTGCCAGCGACGGTGGTCTGCTACCTGAACCAGTGAAGGTGA
CGAAGTCCCGGTGCTGATGGGCGAGCGTTTTGAAGTGGTGGAGGTTAACGATAACAAACCCTTTGACCT
GGTAGCGTCCCGGTGACCCAGATGGGGATGGCGATTGCGCCGTTTGATAAGCCTCATCCGGTAATGCGGGATT
CAGCCGATTGCTATTAGTGCCTCCCGTGGCTTTGCCAGACACATTAAGTAGCCTGCCTGCGTTACCTTCGCTGG
AAGGGCTGACGGTACGCAAGCTGCAACTCTCTATGGACCCGATGCTCGATATGATGGGGATGCAGATGCTAAT
GGAGAAATATGGCGATCAGGCGATGGCCGGGATGGATCACAGCCAGATGATGGGCCATATGGGGCACGGCAAT
ATGAATCATATGAACCACGGCGGGAAAGTTCGATTTCCACCATGCCAACAAAATCAACGGTCAGGCGTTTGATA
TGAACAAGCCGATGTTTGGCGCGGCGAAAGGGCAATACGAACGTTGGGTTATCTCTGGCGTGGGCGACATGAT
GCTGCATCCGTTCCATATCCACGGCACGCAGTTCGGTATCTTGTGAGAAAATGGCAAACCGCCAGCGGCTCAT
CGCGCGGGCTGAAAAGATAACCGTTAAGGTAGAAGTAATGTCAGCGAAGTGTGGTGAAGTTTAAATCACGATG
CACCGAAAGAACATGCTTATATGGCGCACTGCCATCTGCTGGAGCATGAAGATACGGGGATGATGTTAGGGTT
TACGGTAGAGCTCTAG

Amino acid sequence (491 aa): 53.8 kDa (runs at ~55 kDa)

MAERPTLPIPDLLTTDARNRIQLTIGAGQSTFGGKTATTWGYNGNLLGPAVKLQRGKAVTVDIYNQLTEETTL
HWHGLEVPGEVDGGPQGIIPPGGKRSVTLNVDQPAATCWF**HPH**QHKGKTGRQVAMGLAGLVVIEDDEILKMLP
KQWGIDDVPVIVQDKKFSADGQIDYQLDVMTAAVWFGDITLLTNGAIYPQHAAPRGWLRLLRLLNGCNARSLNF
ATSDNRPLYVIASDGGLLPEPVKVSLEPVLMBERFEVLVEVNDNKPFDLVTLPVSQMGMAIAPFDKPHPVMRI
QPIAISASGALPDTLSSLPALPSLEGLTVRKLQLSMDPMLDMMGMQMLMEKYGDQAMAGMDHSQMMGHMGN
MNHMNHGGKFDHFHANKINGQAFDMNKPMAAAKQYERWVISVGDMMML**HPFH**IHGTTQFRILSENGKPPAAH
RAGWKDITVKEGNVSEVLVKFNHDAPKEHAYMA**HCHLLEHEDTGM**MLGFTVEL

HH01-218

Nucleotide sequence (1194 nt):

ATGAATAGCGGGGTAAAAGAGTTCACCTTGGTTGCGGAGCCGGTGGTGC CGCAACTCGCGCCCGGGATGAAAG
CGAACCTGTGGGGCTACAACGGACAATCTCCAGGGCCGACCATCGAAGTGGTGAAGGCGACCGTGTACGCAT
ATTCGTGACCAACAAACTGCCAGAGCACACAAGCGTGCATTTGGCATGGTCAGCGACTGCCGAACGGCATGGAT
GGGGTCACGGGGCTTACTCAGCCAGGTATCGCGCCAGGTAACCTTTGTCTACGAGTTCGTTGCGAAGCGTC
CGGGTACGTTTATGTATCACCCGCATGCCGACGAAATGGCGCAGATGGCGATGGGTATGATGGGCTTTTGGGT
CACTCACCCGAAAAGACCCGTCGCTGCACGCAGTGGACCGGACTTCGTTCTCCTGCTGAACGCCTATGACGTA
GAGCCGGGAAGCTATACGCCGAAAATCAACACGATGCTGGACTTCAATCTGTGGACTTTCAATAGCCGGGTCT
TCCCAGGTATCGACTCGATGCCTGTGCGTCAAGGCGACAAGGTGCGCATCCGAGTTGGCAACCTAACGATGAC
GAATCACCAATCCATTTGCATGGGCACGAGTTCCTTCGTCACCGGCACCGATGGGGGGTGGACACCGCCAGCA
TCGCGCTGGCCCGAGGTCACTACCGACGTCGCGGTGGGGCAGATGCGCGCAGTTGAATTTGTGCCACCGATT
TGGGCGACTGGGCCCTTCCACTGTCACAAGTCCCATCACACCATGAACGCAATGGGGCACAACGTCCCAGCAGT
GATTGGTGTGACCATCGCGGTGTGGCCGAGAAGATTAACAAGATTGTGCCCGACTACATGGTCATGGGCGAC
AAAGGGGGATCGATGGGCGGAATGGAGATGCCGATTCGGGAGAACACCTGCCGATGATGATGGGCGAAGGGC
CATTCGGCGCGGTTGAGATGGGCGGCATGTTTACCGTGTGAAAGTTTCGCAAGGACCAGAAACGCGGCGATTA
CACGGACCCCGCTGGTACAAGCATCCCCTGGGACCGTGGCGTACGAGTGGACCGGCGCGCTGCCGGAGCCT

GAGCGGAGCACGAACGCAGGTGGCCAGTCTATGCCGGCAATGAATAAACCGAATGTCGAGATGACGGTTCGAA
AATCGTCGGGACATGGCGGACACTAA

Amino acid sequence (397 aa): 43.8 kDa

MNSGVKEFHLVAEPVVRELPGMKANLWGYNGQSPGPTIEVVEGDRVRIFVTNKLPEHTSV**HW**HGQRLPNGMD
GVTGLTQPGIAPGKTFVYEFVAKRPGTFMY**HP**HADEMAQMAMGMMGFVWTHPKDPSLHAVDRDFVLLNAYDV
EPGSYTPKINTMLDFNLWTFNSRVFPGIDSMPVRQGDVRIIRVGNLMTN**HP**I**HL**HGHEFFVTGTDGGWTPPA
SRWPEVTTDVAVGQMRAVEFVATDLGDWAF**FCH**K**SH**HTMNAMGHNVPTMIGVDHRGVAEKINKIVPDYVMGMD
KGGSMGMEMPIPENTLPMMSGEGPFGGVEMGGMFTVLKVRKDQKRGDYTDPAWYKHPAGTVAYEWTGALPEP
ERSTNAGGQSMFPAMNKNVEMTVRKSSGHGGH

HH01-482

Nucleotide sequence (1392 nt):

ATGATTTACGCTAGAGACTTCTTCAAGGGCGCGGGCGGGTTCGAGTCAGCGCGGGCCCGCTCAGCAAGGCCG
GCGCGGCCACGCTGCCGGAGGCGATGACCATGGATAACGCCGACACCCATGTGCCGCCGCTGCCGTCCAACGG
CCGTCCGTTCACCCCGGTGGTCACGCTGAACGGTTGGTCGCTGCCGTGGCGGATGAATAACGGCGTCAAGGAA
TTCCACCTGATCGCCGAGCCGGTGGTGCGCAGAAATGGCGCCTGGCATGAAGGCCAATCTGTGGGGCTACAACG
GCCAGTCGCCCCGCCCCGACCATCGAGGTGGTGGAGGGTGACCGCTGCGCATCTTCGTCACCAACAAGCTGCC
GGAGCACACCAGCGTGCCTGGCACGGCCAGCGCTGCCGAACGGCATGGACGGCGTCAACGGCCGTGACGCAG
CCCGGCATCGCTCCCGCAAAAACCTTCGTCTACGAGTTCGTGCGCCAGGCGCCCCGGCACCTTCATGTACCACC
CGCATGCGGATGAGATGACGCAGATGGCGATGGGCATGATGGGCTTCTGGGTACGCACCCCAAGGATCACAC
GCAGTTCGCGGTGGACCGGACTTCGTGTTCTGCTGAACGCCCTACGACATCGACCCGGGCAGCTACACGCCG
AAGACCAACACCATGATGGACTTTAACCTGTGGACTTCAACAGCCGCTCTTCCCCGGCATCGATTCATGC
CGGTGCGGCAGGGCGACAAGGTGCGCATCCGCGTGGCAACCTGACGATGACCAATCATCCGATACATATCCA
CGGCCACGAGTTCTTCGTACCCGGCACCGATGGCGGCTGGACGCCCGCCGCATCGCGCTGGCCGGAAGTGACG
ACCGATATCGCCGTCGGCCAGATGCGGGCGATCGAATTTCGTGCGCCACCGATCCGGGGCAGTGGGCCTTCCACT
GCCACAAAATCGCACCACACGATGAACGCCATGGGCCACAACGTGCCGACCATGATAGGCGTGGACCATCGCGG
CATCGCCGAGAAAGATCAACAAGATCGTGCCGGACTACATGACCATGGGCGACAAGGGCGGCCATGGGCGGC
ATGCAGATGGCGCTGCCGGAGAACACGCTGCCGATGATGTCGGGCGAAGGCCGTTTCGGCGGCGTGGACATGG
GCGGCATGTTACGGTGGTCAAGGTGCGCAAGGAGCAGAAACGCGGCGACTACACGGACCCCGGCTGGTACAA
GCATCCACCGGGCACGCTGGCCTACGAGTGGACCGCGCCTTCCCCGAGCCGGCGCAAGTACACGGCAGGT
GGCCAGTCCATGCCTGCGGTGAACCCGCCAACGTGGAACACGCGTGGCAGAACCAAGCGGCCATAGCGGCC
ACTAA

Amino acid sequence (463 aa): 50.9 kDa (runs at ~48 kDa)

MISRRDFFKAGAVAVSAAAVSKAGAATLPEAMTMDNADTHVPLPSNGRPFNPVVTNLGWSLPWRMNGVKE
FHLIAEPVVREMAPGMKANLWGYNGQSPGPTIEVVEGDRVRIFVTNKLPEHTSV**HW**HGQRLPNGMDGVTGLTQ
PGIAPGKTFVYEFVARRPPTFM**YHP**HADMTQMAMGMMGFVWTHPKDHTQFAVDRDFVLLNAYDIDPGSYTP
KTNTMMDLWTFNSRVFPGIDSMPVRQGDVRIIRVGNLMTN**HP**I**HL**HGHEFFVTGTDGGWTPPASRWPEVT
TDIAVGQMRALIEFVATDPGDWAF**FCH**K**SH**HTMNAMGHNVPTMIGVDHRGIAEKINKIVPDYMTMGDKGGAMGG
MQMALPENTLPMMSGEGPFGGVDMGGMFTVVKVRKEQKRGDYTDPGWYKHPPTLAYEWTGALPEPARKYTAG
GQSMFAVNPPNVELTVRKPSGHS

SM3532

Nucleotide sequence (1869 nt):

ATGTCGCATGATGATTTTCGTGGTCCACGCGGTGGACCGCTGCTGCCTTCGCGGGCGGATTTGTCCAAGGCT
TGGCCTTGGGAGCCGAGTCGCAGGATTAGGTTTCTGGCCCAAAGCCAGTTGGGCGCTCAAGGGGCCGGGACA

ACCCAACGTA CTATCGGGCACTGAGTTTGACCTGACCATTTGGCGAGACGCCGATGAAC TTCACCGGCAAGACC
CGCACCGCGATCACGGTCAACGGGTCCGTTCCGGCGCCGTTGCTGCGGTGGCGGGAAGGCACCACGGTCAACT
TGCGCGTCTCCAATGCATTGCCCGCTAACTCCATCCATGGCGCGGACACCTCCATCCATTGGCACGGCATCAT
TTTGCCGGCCAACATGGACGGCGTGCCGGGTCTGAGCTTTGACGGTATCGGACGTGGTGAGACCTACCCTAC
CGGTTACACCTGCATCAGGGCGGAACCTACTGGTACCACAGCCACTCAGGGTTCAGGAACAAGCCGGGCTCT
ATGGCCGATCGTCATCGACCCATTGGAGCCGGAGCCCTTCAGTTTCGATCGCGACTACGTCGTGATGCTGAG
CGATTGGACAGACCTGGACCCGACGGCCCTGTTTCGATCGTTTGAAGAAGATGCCGGGCCATGACAATTACTAC
AAGCGCACGGTCGGCGATTTTGCGCGCGATGTGAAGCGCAACGGCCGTGTCGGCCACGTTGGAAGATCGCAAGA
TGTGGGGCGTGATGCGGATGACGCCACGGACCTGTCCGACGTCAACGCCAACACCTACACCTACCTGATGAA
CGGCACGACCTCACTGGGCAACTGGACCGGTTTGTTCGCGAGTGGCGAGAAGGTGCGCCTGCGTTTCATCAAT
GGCTCTGCCATGACGTA CTTCGATGTGCGTATTCCGGGGCTGAAGATGACCGTGGTGGCGGCAGATGGCTTGT
ATGTCCATCCGGTTTCCGTCGACGAGTTCGCGATCGCGGTAGCAGAAACCTTCGATGTGATCGTGAGCCCTC
CGGGCAGGACGCATTACCATCTTTGCCAAGACTCCGGTTCGACCCGGCTACATCAGCGGCACGCTCGCTGTG
CGGAAGGATTACGCGCGCCCGTTCGCTCTGTGGATCCCCGGCCGCTGCTGACGATGGCAGACATGGGCATGG
ATCATGGGTGATGGATATGTCTGGCGGCAGCAAGGGCATGGAAGGGCGCTGTGGTGGCGCCATGGGCATGCC
TGGCATGACCCACCTGTCAGCGGTAACGCGACCTCGGCCATGCAGGCCATGCGATGCCCGCCGCGCGCGAT
GGTGCCATGGCAGGCATGCAGCACGGGGCATGCAATCACACCCTGCTAGCGAGACCAACAATCCCTGTTGG
ACAACCAAGCCATGAGCGTGAGTTCGCGCTTGGATGATCCGGGCAATGGCCTGCGCGATAACGGCCGTCATGT
GCTGACGTATTCCATGCTCAAGAGCACCTTTGAAGACCTGACGGACGCGACCCCGGTGCGGAGATCGAGCTG
CATCTGACCGGACACATGGAGAAATCTCCTGGGGCTTCAACGGTCAGAAGTTTTCCGATGTGAGCCGCTGC
GGCTGAACTACGGCGAGCGTATGCGCATCGTATTGGTTAACGACACGATGATGACCCATCCCATCCATTTGCA
CGGCATGTGGAGTGACGTGGAGGACGACAACGGCAACTTCATGGTGGCAAGCACACGGTGGATATGCCGCCA
GGTAGCCGACGCACGTATCGCGTGCGTGCCGATGCGTTGGGCAGCTGGGCGTTCATTTGCCACCTGCTTTATC
ACATGGAAGCCGGAATGATGCGCACGGTGAGGGTCGACGAATGA

Amino acid sequence (622 aa): 68.2 kDa

MSHDDFRGPRGGPLLP SRRRFVQGLALGGAVAGLGFWPKASWALKGPGQPNVLSGTEFDLTIGETPMNFTGKT
RTAITVNGSV PAPI LLRWREGTTVNLRVSNALPANSIHGADTSI **HWHGI** ILPANMDGVPGLSFDGIGRGETYHY
RFTLHQGGTYWY **HSHS**SGFQEQAGLYGPIVIDPLEPEPFSFDRDYVVM LSWTDLDPTALFDR LKKMPGHDNY
KRTVGFDFARDVKRNGLSATLEDRKMWGV MRMPTDLS DVNANTYTYLMNGTTS LGNWTGLFRSGEKVRLRFIN
GSAMTYFDVRI PGLKMTVVAADGLYVHPVSVDEFRIAVAETFDVIVEPSGQDAFTIFAQDSGRGTGYISGTLAV
REGLRAPVPSVDPRPLLTMDMGM DHGSMDSGGSKGMEGGCAAMGMPGMT PPVSGNATSAHAGHAMP AAGD
GAMAGMQHGGMQSHPASETN NPLLDNQAMS VSSRLDDPGNGLRDNGRHVLTYSMLKSTFEDPDGRDPGREIEL
HLTGHEKFSWGFNGQKFS DVEPLRLNYGERM RIVLVNDTMMT **HP IHLHGM** WSDVEDDNGNFMVRKHTVDMPP
GSRRTYRVRADALGSWAF **HCHLLYHMEAGM** MRTVRVDE

SM4696

Nucleotide sequence (1809 nt):

ATGAATACCCGCAATCTCCCGGTCCGGGCGCTGTGCCATGCCGTCGCGCCGCTGTTTCGTGCAGGGCCTGG
CCGCCGGTGGCGTGGTTCGCCGCATCGCCGCTGTGCGCGTGCCGCAACGCGCGCTCGCCGCCGCACTGCCGC
CCCACGGCTGGCCGGTGC CCCC GCGTGTCTCAGCGATACCCGCATCGA ACTGGCCATCGGCGAATCGCTGGCC
AATTTACCCGGCCGACCCGCCGGCGATCACCGTCAACGGCTCGTTGCCGGCCCCGATCCTGCGCTGGCGCG
AAGGCCAGACCGTGACCTGTTTCGTGCGCAACACGCTGGAACGCCACCCGACCTCCATCCACTGGCACGGCAT
CCTGCTGCCGGCCAACATGGACGGCGTGCCGGGCTGAGCTTCAACGGCATCGGCCCGGCGAAACCTACCAC
TACCACTTCCAGCTGAAGCAGTCGGGCACCTACTGGTACCACAGCCACTCGATGTTCCAGGAACAGGCCGGCC
TGTACGGCGCGCTGATCATCGACCCGGCCGAACCGGCGCCGTACCACCACGACCCGCGAGCACGTGATCATGCT
GTCGGACTGGACCGACATGGATCCGGGCGCGCTGTTCCGGCGCATGAAGAAGCTGGCCGAGCACGACA ACTAC
TACAAGCGCACCTGCCGACTTCTTGC GCGACGCGAAACGCGACGGTTGGTCCGCGCGCTGTCCGACCGTG
GCATGTGGGCGCGGATGCGGATGACGCCACCGGATATCCGATATCAACGCGCATACCTACCTACCTGAT
GAACGGCACCCGCCCGGCAACTGGACCGCCGCTGTTCGCGAGTGGCGAAAAGGTGCTGCTGCGCTTCATC
AACGGCGGGTTCGATGACCTACTTCGACGTGCGCATCCCGGTTTGAAGATGACCGTGGTGGCCGCTGACGGCC
AGTACATCCACCCGGT CAGCATCGACGAATTCGCGATCGCTCCCGCCGAAACCTACGACGTGCTGGT CGAACC
CACCGGCCAGGATGCGTTACCATTTTCTGCCAGGACATGGGTGCGACCCGGCTACGCCGCCGGCACGCTGGCG
GTACGCTACGGGCTGCAGGCACCGATTC CCGCGCGCGACCCACGCCCGTTGCTGACGATGAGCGACATGGGGC
ACGACATGGGCGGTGGTGGCCACGGCGCCACGACATGGCCGCGATGAAGAGCACGGAAGGTAGCTGCGGCGC

CAGCATGGGCCACGGTGCACGCGCGGGTCCCGCCAGCAAGGTACCGAAGCATCCGGCCAGCGAGCGCAAC
AACCCGCTGGTGGACATGCAGAGTTCGGCCACCGAGCCGAAGCTGGACGACCCCTGGCATCGGCCTGCGCGACA
ACGGCCGCCAGGTGCTGACCTACGGCGCGATGCGCAGCCTGTTCGAAGACCCGGATGGCCGCGAGCCGAGCCG
CGAGATCGAGCTGCACCTGACCGGCCACATGGAGAAGTTCCTCTGGTTCGATGGCGTCCCCTTCGCCAGC
GCCGAGCCGCTGCGGCTGAACTACGGCGAGCGCATGCGCATCGTGTGGTCAACGACACCATGATGCAGCACC
CGATCCACCTGCACGGCGTGTGGAGCGATCTGGAAAACGCGCAAGGTGAGTTCAGGTCCGCAAGCACACCAT
CGACATGCCGCCGGTACCCGCCGAGCTATCGCGTGCAGCGCGATGCACCTCGGCCGCTGGGCCTACCACTGC
CATCTGCTGTACCACATGGAAGCGGGCATGATGCGCGAAGTGAGGATCGAAGAATGA

Amino acid sequence (602 aa): 66.3 kDa

MNTRNLPFGAVPMPSSRRLFVQGLAAGGVVAGIAAVGVPQRALAAATAAPRLAGAPAVLSDTRIELAIGESLA
NFTGRTRPAITVNGSLPAPILRWREGQTVDLFVRNLTLEHPTS **IHW**HGILLPANMDGVPGLSFNGIGPGETYH
YHFQLKQSGTYWY **HSH**SMFQEQAGLYGALIIDPAEPAPYHHDREHVIMLSDWTDMDPGALFRMKKLAHDNY
YKRTLPDFLRDAKRDGWSAALS DRGMWGRMRMPTDISDINAHTYTYLMNGTAPAGNWTGLFRS GEKVLLRFI
NGGSMTYFDVRI PGLKMTVVAADGQYIHPVSI DEFRIPAETYDVLVEPTGQDAFTIFCQDMGRTGYAAGTLA
VRYGLQAPI PARDPRPLLTMSDMGHDMGGGGHGGHDMAMKSTEGSCGASMGHGAHGGGAASKVPKHPASERN
NPLVDMQSSATEPKLDDPGI GLRDNGRQVLTYGAMRSLFEDPDGREPSREIELHLTGHMEKFSWSFDGVPFAS
AEPLRLNYGERMRIVLVNDTMMQ **HP** **IHL**HGVWSDLENAQGEFQVRKHTIDMPPGTRRSYRVRADALGRWAY **HC**
HLLYHMEAGMMREVRIEE

NGR147

Nucleotide sequence (1401 nt):

ATGACGCCGTTGCTCAGCCGCCCACTTTTCTCCAGGGCTCGGCCGCACTTGGCGGTGCCTTTGCGCTTGGCG
CCGGGCTCGCCGACGCGCCGGGGCGGTTCCCGAAGCTCAGCTCCTCACGGCGCAGCTGGCAAAGGCGCAGAT
TGCCGCCGATGGTGTACGCCCCGTCATGACATACGGACTTGGCGACGCGGTCCATACCCACCTGCCGCCCGTC
CTGAGGATGCGCAAGGGCGAACCCTATGCGGGCAAGGCTCGTCAACAGGCTCGACGAGCCGACGACGGTCCATT
GGCACGGCTTGGCGATCGCCAATGCGATGGACGGCGTGC CGGAGATGACCCAGGCTTACGCTATCCCGGCGA
TCACTTCGACTACAGTTTCACCCCGCCGGACGCCGGCACGTTCTGGTACCATCCCCATTGCAACACGCTGACG
CAGATGGGCCATGGCTGACCGGGGTGATCGTCTCGAAAACCCGGCGGACCCGACGTTTCGACGCGGAGATCG
TCCTCAATCTTCGTGACTGGCGGTGGGCGCCGGGGGAAATACATCGACCCCTTCAAGCCACGCGATGCCGC
GCGCGCGGCACCTACGGGACCGTCAGAACGGCGAATTGGCAGCAGGAACCTGTCTATGACGACCCGGCCGGC
GGGCTCGTTCGCTCAGGATCGCCGCGACCGACGTGACGCGCATCTACACGATCGGCCCTCGAGGGGGCCGAGG
CGAAGTGGTCCGCTCGACGGCAATCCGGTCCGATCCTCCGTTCCGTCGACCGGCTCGACATCGGGCCGGG
ACAGCGCGTCGATCTCATCATGCGCATGCCCGACAGCGAAGGGGCAGGCGGACACTGGGCAACTCCCGCGC
TCGAACCCGTGGACGATCGCGACGCTCAGGTCAGTCCGCCCTTCCCTAAAGCGGGACCTTGGCGACGTCGCC
CCCTTCCGGCCAATCCGGTTCGCTAGAGCCGACCTGTGACGGCGCAGCGGATTCCCCTCGAACTCACGGCAAC
GGCCGAACACACGGCGGTACCATCGGTCTGCGGTTCGCTCGGTTACACCTTTTGGGCGCTCAACAAGGTTCCG
TGGCCGGGGGACACGCCCGACCCGATTGCACCGATCGAGGAACGAAGCTCGGCAAAAGCTATGTGCTCGAGA
TCGCCAACCGCACCCCGCACGCGCACCCGATCCACCTGCATGGCTTGGATTTCCGCATCCTGAGCTCGAACAA
GCGGAGCTTTCTGCCGCCACCGACCGACACGATCCTGCTTTTGCCTGACGAGCAGGCGGAGGTGGCGCTCGTT
GCCGACAATCTCGGCGACTGGGTGATCCACTGCCACATCATCGAACATCAAAGACCCGGAATGACCGGCTATT
TCAGAATTATCTGA

Amino acid sequence (466 aa): 50.3 kDa

MTPLLSRRTFLQGSAAALGGAFALGAGLAARAGAVPEAQLLTAQLAKAQIAADGVTPVMTYGLGDAV **HT**HLPPV
LRMRKGEPIYAARLVNRLDEPTTV **HW**HGLRIANAMDGVP EMTQAYVYPGDHFDYSFTPPDAGTFWY **HP**HCNTLT
QMGHGLTGVI VVENPADPTF DAEIVLNLRDWRGAGGKYIDPFKPRDAARGGTYGTVRTANWQQEPVYDAPAG
GLVVRVRIAATDVTRIYTI GLEGAEAKVVALDGNPVDHPFLDLRLDIGPQRVDLIMRMPDSEGAGATLGNFRG
SNPWTIATLRSVGPLKRDLDGVA PLPANPVARADLSTAQRIPLELTATAEHTAVPSVCGSLGYTFWALNKVP
WPGDTPDPIAPIEELKLGKSYVLEIANRTPHA **HP** **IHL**HGLSFRILSSNKRSFLPPPTDITLLLPDEQAEVALV
ADNLGDWVI **HC** **I**EHQKT **GM**TGYFRII

NGR578

Nucleotide sequence (1350 nt):

ATGTTCAACAGAAGACAGTTGCTCGGCGGAGCGCCGCATTGGTGTGCGACCGCCGCTGGGCGAAGACTTCCA
ACATGGGTCTGCCGAGGCCCGCTGATGGAGAAGGCGGAGACGCAGCCCGCGTAAAGCCGACGTCCGGTCC
GGACTACAACCCCTGTTGTTACCCCTAACGGCTGGACACTGCCCTACCGGATGAACAACGGCGTCAAGGAATTC
CACCTCGTCGCCGAACCGGTCGAGCGGAAATGGCGGAGGGCATGACCGCTATCTCTGGGGCTATAACGGCC
AGTCGCCGGGTCCGACCATCGAAGCGGTCGAAGGCGACCGGGTGCGCATCTTCGTACGAACAAGCTGCCGGA
GCACACGACGATCCACTGGCACGGCATGATCCTTCCCTCCGGCATGGACGGCGTCGGCGGCCCTGACGCAGCCG
CACATCCCGGTCCGCAAGACCTTCGTCTACGAGTTCGATCTCGTGAAGTCCGGCACCTTCATGTACCACCCGC
ATTCCGACGAGATGGTGCAGATGGCCATGGGAATGATGGGCTTCTTCGTCTCCATCCGAAGGACCCGAAGTT
CATGTCCGTGCGACCGCATTTTCGTCTTCTGCTCAACGCCACGACATCGACCCCGGCTCCTACGTGCCGAAG
GTCATGGAGATGACCGACTTCAACATGTGGTGTGGAACAGCCGGGTGTTCCCGGACATCAGCCCGCTCGTCG
TGTC AAGAACGACCGGGTCCGGGTCCGGGTCCGCAATCTCAGATGACCAACCACCCGATCCACATGCACGG
CTACGACTTCGAGTACCTGCACCGACGGCGGCTGGGTGCGCGCGGAGGCGCTGGCCGGAGGTGACGATC
GACATCCCGGTCCGAGCGATGCGGGCTATGAGTTCGACGCCAAATACGCCGGCGACTGGGCGATCCACTGCC
ATAAGTCGCACCACACCATGAACGCCATGGGACATGACATCCCGACCTTCATCGGCGTCGACAAGACGAAGGT
CGCCGAGAAGATCAAGAAGATCCGCCCGGAATACATGCCGATGGGCACCAAGGGCATGGCCGACATGGGCGAG
ATGGAGATGGAGATCCCCGAAAAACACCATCCCGATGATGACCGGTGGGGTCCGCACGGCCCGATCGAGATGG
GCGGCATGTTCTCGGTCTCAAGGTGCGCGAAGGCATCTCGGCCGGCGATTACGCCGATCCCGGCTGGTACGA
AAACCCACCCGTTACCCAGGCCTGGGAGTGGACGGGCGAAGTTCGCCGACGCGACCAAGTCCAAAGACGCCAAG
ACCCAGATCACGCCGAAACACTCAAGCCACGGCTGA

Amino acid sequence (449 aa): 50.2 kDa

MFNRRQLLGASAALVSTAAWAKTSNMGLPEAAVMEKAETQPPVKPTSGPDYNPVVTLNGWTLPYRMNNGVKEF
HLVAEPVEREMAEGMTAYLWGYNGQSPGPTIEAVEGDRVRI FVTNKLPEHTT **IHW**HGMILPSGMDGVGGLTQP
HIPVGTKFVYEFDLVKSQTFMY **HPH**SDEMVMQAMGMMGFFVHPKDPKFMVDRDFVFLNAYDIDPGSYVPK
VMEMTDFNMWCWNSRVFPDISPLVVSKNDRVRVRVGNLMTN **HP** **IHM**HGYDFEVTCTDGGWVRPEARWPEVSI
DIPVGAMRAYEFDAKYAGDWA **IHCHKSHHTM**NAMGHDIPTFIGVDKTKVAEKIKKIRPEYMPMGTKGMADMGE
MEMEIPENTIPMMTGWGPHGPIEMGGMFSVVKVREGISAGDYADPGWYENPPGTQAWIEWTGEVDPATKSKDAK
TQITPKHSSHG

NGR600

Nucleotide sequence (1491 nt):

ATGCCAGACTTCTCACGTCGTCAGTTTCTCACTTCTGCGCCGCCCGCGGTATGGCTGTAGCCAGTCCGTCAT
TGCTGGTCCCCGCGGCCACGCTTCCGCAACGCCGCAACCTTGATCAGGGCCGGAACCAGGACGATAGAGGT
CAACGCCCGGGCCGCCACGGTGTACGGTCTGATGCAGGCTAACGGCACGCACGGCTTATCGCGGAGGCGC
GGTCCGTTCCAAGTCCGGTTGGAAAACGGGCTCGACGAGGAGACGCTGATTCATTGGCACGGCCTTACCCCTC
CGTGGAAGCAAGACGGTGTTCGAATGTCTCTCAGGCGCTGCTTTCGCCCGGGCGCGCACAACACTACGACTT
CCTGCTGAACCGGGCCGGCACCAACTGGATGCACTCCCATCACGGTCTGCAGGAACAGCGCCTGCTGGCGGCG
CCGCTTGTCGTGCGCGGCCCTGCCGAAGCTCGCGAAGACGTGCAGGAAGTGGTTCCTCTTCCACGATTTCA
CGTTCGCGATCCGGCTGAGGTGCTGGCCGGCTGCAGGGTGTGAACAGGTGCCCAGGCACCGGCCGCGCC
GGACCATGCCCAATGGGCCACGCAGCGATGGGTTCATGGTGTCCGGCGCCCGCCGAGGCGGGCATTCGGC
GGAAGGGGCGTGGCTGCGCCCGCAGGTGCGCACCTGAACGATGTGCAATACGACGCCTTCTCGCCAATGACC
GTACGCTTGACGATCCCGAAATCTTCCGGTGGAGCGCGGTGGGCGGGTGCCTGCGGCTGATCAACGGCGC
TGCTGCGACAAACTTCTGGATCGATCTTGGCGGGCTCGAGGGGCGCTGATCGCGGTCGACGGAATGCCGGTC
GAGCCGGTCCGCGACGGCTTTCGAATTCGCCATCGCGCAACGCTCGATATCGCGCTGGACCTGCCGAAGA
ACGAAGGCGCATGGCCGATCTTTGCCATCCGCGAGGAGAGCGTGC GCGCACCGGTTTCATACTGGCCACCGC
CTCCGGTGCCGTAGAGCGCGTCGGCGGAGATGCAGACGAAGCCGCCGCGCCGTAGGGCTCGCGCTTGAAGGA
AAATTACGGGCTGCGCAGCCGCTGGCGGGACGTCCGGCCGACCGCGAGCACACGGTACCATACCCGAGGCAC
CCGGTTATGTCTGGATGCTGAACGGCAAGAGGCACGGTGAAGATGCGCCGCTCCCGGTGGCCGAGGGCGAGCG
CGTCGAGATCACCTTCGTGACCACACCACGATGGCGCACCCGATGCATCTGCACGGCCATCACTTTCAGGTG

GTTGCCATCAACGGCAAGCGGTTCCCGGGCGCGGTGCGCGACACCGTCTCGTGCCGTCGCAGGGTAGCGTGA
CAATCGCTTTCGATGCGGTGAACCCAGGCAAGTGGGCGCTTCACTGCCACCACCTCTACCATATGAACGGAGG
AATGATGACGGCGGTCGAGTATCAAACCTAG

Amino acid sequence (496 aa): 53 kDa

MPDFSRRQFLTSCAAAAGMAVASPSLLVPARHASATPATLIRAGTRTIEVNGRAATVYGLMQANGTHGLIAEAG
GPFQVRL ENGLDEETLI **HW**HGLTPPWKQDGVNVSQALLSPGGAHNYDFLLNRAGTNW**MS**HHLQEQRL LAA
PLVVRGPAEAREADVQEVVVLFDFTFRDPAEVLQGLQVEQVAQAPAAPDHAAMGHAAMGHGAPAPAAGGHS
GRGVAAPAGAH LNDVEYDAFLANDRTLDDPEIFRVERGGRVRLRLINGAAATNFWIDLGGLEGR LIAVDGMPV
EPVRGRRFEFAIAQRLDIALDLPKNEGAWPIFAIREGERARTGFILATASGAVERVGGDADEAAAAVGLALEG
KLRAAQPLAGRPADREHTVTITEAPGYVWMLNGKRHGEDAPLPVAEGERVEITFRDHTTMA**HPMHL**HGHHFQV
VAINGKRFPGAVRDTVLVPSQGSVTIAFDVNP GKWAL**HCH**HL**YHM**NGG**MM**TAVEYQT

NGR688

Nucleotide sequence (1350 nt):

ATGTTCAACAGAAGACAGATACTGGGAGCAGGAGCGGCGCTTGTGTCTCGGCCGCTGGGCGAAGACCTCCA
ACACGGGACTGCCCAGGCTGCGGTGATGGAGACTGCGGAGACGCAGCCGCCGCTCAAGCCGACCTATGGCCC
GGACTATAACCCCGTCGTCACGCTCAACGGCTGGACGCTGCCCCACCGGATGAACAACGGCGTCAAGGAATTC
CACCTCGTCGCCGAGCCGGTCGAGCGCGAGATGGCGGAAGGCATGACCGCTATCTCTGGGGCTACAACGGCC
AGTCGCCGGGACCGACGATCGAGGCCGTCGAGGGCGACCGGGTGC GCATCTTCGTACCAACAAGCTGCCCGA
GCACACGACCATCCACTGGCACGGCATGATCTGCCGTCCGGCATGGACGGCGTCGGCGGCCGTGACGCAGCCG
CACATCCCGGTCGGCAAGACCTTCGTCTACGAGTTCGACCTCGTGAAGTCGGGCACCTTCATGTACCACCCGC
ACTCCGACGAGATGGTGCAGATGGCGATGGGGATGATGGGAATGTTTCATCATCCATCCCAAGGACCCGAAGTT
CATGCCGGTCGACCGGACTTCGTCTTCTGCTCAACGCCTACGACATCGATCCCGGCTCCTACGTCCCGAAG
ATCATGGAGATGACCGACTTCAACATGTGGTGTGGAAACAGCCGCTGTTTCCCGACATCAGCCCGCTGGTGG
TTTTCGAAGAACGACCGGGTGC GCGTCCGGGTCCGGCAACCTGACCATGACCAACCATCCGATCCACATGCATGG
CTATGACTTCGAGGTCACCTGCACCGATGGCGGTTGGGTCCGGCCGGAGGCCCGCTGGCCAGAGGTCAGCATC
GACATCCCGGTCGGGGCGATGCGGGCCTACGAGTTCGACGCCAAGTACGTGGCGACTGGGCGATCCACTGCC
ACAAGTCGCACCACACGATGAACGCCATGGGACACGACATCCCGACCTTCATCGGCGTCGACAAGAAGACGGT
CGCCGAGAAGATCAAAAAGCTTCGCCCGGAATACATGCCGATGGGGACCAAGGCATGGCCGACATGGGCGAA
ATGAAAATGAAAATCCAGAGAACACCGTGCCGATGATGACCGGCTGGGGACCGCACGGCCCGATCGAAATGG
GCGGCATGTTCTCGGTGTCAGGTTTCGCGAAGGCATCTCGGCCGGCGATTACGCCGATCCCGGCTGGTACGA
AAACCCACCCGGAACCCAGGCCGGGAGTGGACGGGCGAGCTCCCCGACGCAACCAAGGCCAAAGACGCCAAG
ACCCAGATCACGCCGAAACATGGAAATCACGGCTGA

Amino acid sequence (449 aa): 50.2 kDa

MFNRRQILGAGAALVSSAAWAKTSNTGLPEAAVMETAETQPPLKPTYGPDYNPVVTLNGWTLPHRMNNGVKEF
HLVAEPVEREMAEGMTAYLWGYNGQSPGPTIEAVEGDRVRI FVTNKLPEHTTI **HW**HGMILPSGMDGVGGLTQP
HIPVGKTFVYEFDLVKSGTFMY **HP**HSDEMVMQAMGMMGMFI IHPKDKFMPVDRDFVFLNAYDIDPGSYVPK
IMEMTDFNMWCWNSRVFPDISPLVVSKNDRVRVRVGNLMTN**HP**I**HM**HGYDFEVTCTDGGWVRPEARWPEVSI
DIPVGAMRAYEFDKAYVGDWA I**HCH**K**SH**HT**MN**AMGHDIPTFIGVDKKTVAEKIKKLRPEYMPMGTKGMADMGE
MEMEIPENTVPMMTGWGPHGPIEMGGMFSVVKVREGISAGDYADPGWYENPPGTQAWIEWTGLP DATKAKDAK
TQITPKHGNHG

Ips1282

Nucleotide sequence (1665 nt):

ATGTTACGTCGCGATTTTTTGAATAAAGTGCTATCACATATGCTGCAAGTGCCTATTTGGAGCCGCA
TCGCGGTAGCCGCTGAAAACATATCCACATTTGGCTATCCCGCCGCTATTAGAGCCTGATGCGCAAGGCAACAT
TCAATTGGCGATTAAGCAAGGCACATCCCAATTTGTTCCCGGTAAAAAACCACCACGTGGGGATATAACGGT
GATTTACTCGCCCTGCGATTAGCTTAAAAAATGGCCAGAGTGTAAATGTCAATATTGTGAATCAGTTGGCAG
AAGATAACCACCGTTCCTGCGATGGCTTAGAAAATTTAGGCGAACAGGATGGCGGGCCACAGGCAATTATTCA
ACCTGGTGCAACACGCAAAAGTCAATTTTACCATTGATCAAAGTGAAGCAACCTGTTGGTTCCACCCTCATACT
CATGGAAAAACCGGCTACCAAGTCGCTATGGGGCTGGCGGGGTTAGTCTGATTAAGATGAAAATAGCGAAA
AACACGGTTTGCCTAAGTCATTTGGGGGTTGATGATATCCCTGTAATTTTACAAGATAAGCGTTTAAAAGATGA
CGGGCAAAATCGATTATCAATTAGATGTAATGAGTGTGCGGTAGGCTGGTTTGGTGATTTTCATGCTGACTAAC
GGCGCGGTTTACCCGAAACATGTTGCGCCAAAAGGTTGGATCCGCTTGGCATTACTCAATGGTTGTAACGCAC
GTAGTTTGAATATTTCAACCAGCGATGGTTCGTAATAATGTACGTTGTTGCCAGTGTGGCGGCTTACTTGTCTGA
ACCGGTTGCATTGACTGAATTACCGATTTTAAATGGGTGAGCGTTTGTAGGTATTAATTGATGCTTCAGATGGT
CATGCGTTTGTATTAGTCACTTTACCTGTAGGGCAAATGGGGATGACTATCGCACCATTTGACCAGCCTCTGC
CTGTTTTACGTATCGAAACCACATTAACAAAATGGCGAAGGAAAAATTACCTTCAACTCTTGCTGCAATCTCTGC
AATCCCTCTTTAGCAGGGTTAAATCGCCGTCGTTTCCATTTGATGATGGATATGCGTTTAGATATGCAAGGC
ATGATGATGTTACGTGAAAAATATGGCGCACAAAGCGATGGGGAACATGGGTGGTTCATGGCGCGATGATGGGTA
ATATGGATCATGGCGCCATGATGAACAGTGGAAACATGGATCATGGGAATATGAGCCAAGGCAAAATGAACCA
CGGTAACATGAGTGGCGGCAATATGCGTGGTGGCATGGGTAATGGCGGGCCTGCGGTACTGGTGTGGTGAT
TTAGATATTCATAACGCCAATACTATCAACGGGCAAGTGTTTTCAATGACGCATCCAGCGTTTGTATGCGCCAA
TGAACCGCCAAGAAATTTGGGTGGTTTCAAGGGCGTGGAGACATGATGTTACACCATTCCACGTACACGGAAC
TCGCTTTCGTATTTTAAAAGAAAAATGGCCAAGCCGTTGAGCCGCATCGCCAAGGCTGGAAAGATATTGTCAA
GTTGAAGGGCAGGTGAGCGAAAATTTAGTGGAAATTTAAACACCCTGCGACGAAAGAACACCCCTATATGGCGC
ATTGCCATTTATTAGAACACGAAGACACAGGAATGATGCTCGGCTTTACAGTGAGTTAA

Amino acid sequence (554 aa): 60.3 kDa (runs at ~65 kDa)

MLRRDFLKLSAITYAASALPIWSRIAVAENYPTLAI PPLLEPDAQNIQLAIKQGTSQFVPGKKTITWGYNG
DLLGPAISLKNQSVNVNIVNQLAEDTTVHWHGLEISGEQDGGPQAI IQPGATRKVNFTIDQSEATCWFHPHT
HGKTGYQVAMGLAGLVLIKDENSEKHGLPSHWGVDDIPVILQDKRLKDDGQIDYQLDVMSSAAVWGFDFMLTN
GAVYPKHVAPKGIWRLLRLLNGCNARSLNISTSDGRKMYVVASDGLLAE PVALTELPILMGERFEVLIDASDG
HAFDLVTLVPVGMGMTIAPFDQPLPVLRIETTLTNGEGKLPSTLAAISAI PSLAGLNRFRFHLMMDMRLDMQG
MMLLREKYGAQAMGNMGHGHAMMGNMDHGMNMSGTMDHGNMSQGMNHGNMSSGNMRGGMGNGGHCCTGAGD
LDIHNANTINGQVFSMTHPAFDAPMNRQELWVVSGRGDMMLHPFHVHGTRFRI LKENGQAVEPHRQGWKDIVK
VEGQVSEILVFEFKHPATKEHPYMAHCHLLEHEDTGMMLGFTVS

Optimized nucleotide sequence for expression in *E. coli*:

GCTAGCATGTTACGCCGTGACTTCCTTAAACTGTCGGCGATTACCTATGCTGCCAGCGCCTTGCCCATTTGGT
CGCGCATCGCTGTTGCAGCGGAAAACTACCCGACATTAGCGATTCCGCCACTGCTGGAGCCGGATGCGCAAGG
GAACATTCAGCTGGCAATCAAACAGGGCACCAGTCAAGTTGTTCCGGGGAAGAAAACCCTACGTGGGGCTAT
AATGGCGATTTGTTGGGCCCTGCCATCTCCCTGAAGAATGGACAGTCTGTAAACGTGAACATTTGTAACCAAC
TGGCAGAGGATAACCACCGTGCATTGGCAGGACTCGAGATCAGTGGCGAACAAGATGGCGGGCCCTCAAGCCAT
CATTACGCCAGGAGCCACCCGTAAAGTGAACCTTACCATTGATCAGAGCGAAGCAACGTGCTGGTTTCATCCC
CATACTCATGGTAAAACGGGTTATCAGGTTGCCATGGGTCTGGCGGGCCTTGTGTTGATCAAAGATGAAAAC
CAGAGAAAACCGCCTGCCAAGTCATTTGGGGCGTGGATGACATTCGGGTCACTCCTGCAGGACAAACGCCTGAA
AGATGATGGTCAAAATGACTACCAGTTAGATGTCATGTCCGCTGCAGTGGGCTGGTTTGGCGACTTCATGCTG
ACTAACGGTGCAGTCTATCCGAAACATGTGGCACCAGAAAGGCTGGATTCCGGTTACGCCTCCTGAACGGTTGCA
ATGCGCGCAGCTTGAACATCAGCACCTCAGATGGGCGGAAAATGTACGTAGTTGCTTCGGATGGCGGGTTACT
GGCAGAACAGTGGCCCTTACCGAACTGCCGATTTCTGATGGGAGAACGCTTCGAAGTGTCTATTGATGCATCT
GACGGTCACTGCTTTGACCTGGTTACACTGCCTTAGTTCAGTGGGCATGACGATTCACAGCTTCGATCAGC
CGTTACCGTTCCTGCGCATTGAAACCCTCTGACCAATGGCGAAGGTAAACTGCCCTACACTTGCAGGCCAT
TTCTGCGATTCCGCTGACTGGCTGAAACCGTGAACCGTCTGCTTTTACCTGATGATGGATATGCGCTTAGACATG
CAGGGTATGATGATGCTGCGTGAAAAGTACGGTGCAGCAAGCTATGGGCAATATGGCGGGTTCATGGAGCCATGA
TGGGGAATATGGATCATGGGGCGATGATGAATAGTGGCACGATGGACCACGGGAACATGAGCCAGGGTAAAAT
GAATCACGGCAACATGTCGGCGGCAATATGCGCGGTGGTATGGGTAATGGTGGGCATTTGTGGTACAGGCGCC
GGTATCTCGACATCCCAATGCGAACACGATCAATGGCCAAGTCTTTAGCATGACGCACCCTGCCTTTGATG

CGCCCATGAACCGTCAGGAAATCTGGGTTGTATCCGGGCGTGGGGACATGATGCTGCATCCGTTTCATGTTCA
CGGTACTCGCTTTCGTATCTTGAAGGAAAATGGTCAAGCGGTGAACCGCATCGCCAGGGATGGAAGGACATC
GTGAAAGTCGAAGGACAGGTTTCGGAGATTCTGGTCGAGTTCAAACATCCAGCGACCAAAGAGCACCCGTATA
TGGCTCACTGTCACCTCCTGGAACATGAAGATACGGGTATGATGCTTGGCTTACCCTGAGCGAGCTC

Ips2204

Nucleotide sequence (1629 nt):

ATGCAACGTCGTGATTTTTATTTAAATTGAGTGCCGCGATGGGGCGCTGCCAGTGCCTGCCATTATGGAGTCGTT
CGCTGATGGCGGAGCAGGTGCGTCTGGAGCTTCCCGTTCCGGCGTTATTAACGGCGGATCCGCGCGGGATAAT
CAATATTGCCGTGCAGCAGGGACAGACTCAATGGATGGGAAAATCCGTTACCCTTGGGGTTATAACGGTAAT
TTGCTGGGGCCGGCGATTCAACTGGATCGCGGTAAACCGGTGACGCTGAATCTGCATAACACATTGCCGGAAG
CGACCACTATTCCTGACGCGGTCTCGCCCTGCCCGGTGAGGTAGACGGTGGTCCGCAGGCAGTGATTGCGCC
AGGAGCCTTCCGCGTGTGAGCTTTACGCCTGATCAACCCGGCGGACTTGTGGTTCCATCCCCATCAGCAT
GGCCGACCGGTTATCAGGTGGCACAGGGATTGGCCGGGCTGGTTATCCTGAAAGATGATGCGGGCGAGAAAT
TACTGTTGCCATAAAATCTGGGGGGTGGATGATATCCCGTTATCCTGCAGGATAAACGGCTGAGTACTGATGG
CAGTAAGATTGATTATGCGCTGGATATGATGAGTGCAGCGGTGGGGTGGTTTGGCAATACCATGCTGACTAAC
GGGGTGATTTATCCGCTGCAAGCGGTGCCGCGGGGTTGGTTACGCCTGCGTCTGCTGAATGGTTGTAATGCC
GATCGCTGAATCTGACCACAGTGACCAGCGTCCCTGTATGTGATCGGCAGTGATGGCGGGCTGCTGGCGGA
ACCGGTGAAGCTGACTGAACTGTGATGATGCCAGGCGAACGCTTTGAAGTGTGGTAGATACCCACGATGGC
AAAACCTTTGACCTGCAGACGCTGCCAGTTCGGCAGGCGGGCATGACGTTAGCGCCGTTTGACCAGCCGTTGC
CGCTGCTAACTATTCAGCCGTTGCAGATCCTCGCCAGTGGGTCTCTGCCGTGATAAGCTGGTGGATATGCCAGC
ATTGCCAGCACAGATGGGGTTAAAGAGCGTTGGTTGCAACTCATGATGGATAACCGAATGACGATCGCGGT
ATGCAGGCTCTGATGGAAAAATATGGTTCGTCAGCGCTGGCGGGGAGCCGCCACGATGCACATAACACTGACG
GTGATATGTCTACGCGGTTGGCAAGGGAAATGGATCATGACGCGATGGCGGGCATGGCTAACGGTTCCATGCC
CGGTATGGCACACAATCTGCCCGCTCAGAAACCATATGATTTCCATCATGGCAATAAAATTAATGGGGTTGCC
TTTGATATGAACACACCTTCGTTTCGACGTGAAACAAGGTGCGCTTGAGAAGTGGACCATTCTGGCGAGGGTG
ACGGCATGCTACATCCGTTCCATATTCACGGCGCTCAGTTTCGTATCCTTTCCGAAAATGGTAAACCCGCCGC
TGCCCATCGCAGTGGCTGAAAAGACATGGTTTCGGGTCGAGGGCTGGCGTAGCGAGGTGCTGGTACGTTTTAAT
CATCTTGCACCAAAGACCAGGCCTATATGGTCACTGCCATCTGCTCGAGCATGAGGATACCGGGATGATGC
TGGGGTTACGGTGAGCGCTTGA

Amino acid sequence (542 aa): 59.2 kDa (runs at ~60 kDa)

MQRDRFIKLSAAMGAASALPLWSRSLMAEQVRLLELPLVALLTADPRGI INIAVQQGQTQWMGKSVTTWGYNGN
LLGPAIQLDGRKPVTLNLHNTLPEATTIHWHLGLALPGEVDGGPQAVIAPGAFRRVVSFTPDQPGATCWFHPHQH
GRTGYQVAQGLAGLVILKDDAGEKLLLPKIWDVDDIPVILQDKRLSTDGSKIDYALDMMSATVGFNTMLTN
GVYIPLQAVPRGWLRLRLLLNGCNARSLNLTSDQRPLYVIGSDGLLAEPVKLTELSSMPGERFEVLVDTHDG
KTFDLQTLFVRQAGMTLAPFDQPLPLLTIQPLQILASGSLPDKLVDMPALPAHDGVKERWLQLMMDTELDLDRG
MQALMEKYGRSALAGSRHDAHNTDGDMSAVGKGMHDAMAGMANGSMPGMAHNLPAQKPYDFHHGNKINGVA
FDMNTPSFDVKQGALEKWTISGEGDMLHPFHIIHGAQFRILSENGKPAAAHRSGWKDMVRVEGWRSEVLVRFN
HLATKDQAYMAHCHLLEHEDTGMMLGFTVSA

Optimized nucleotide sequence for expression in *E. coli*:

GCTAGCATGCAACGTCGCGATTTTCATCAAGCTGAGCGCAGCGATGGGTGCGGCCCTCGGCCCTGCCCTCTGTGGT
CACGTAGCCTTATGGCAGAACAAGTGCCTCTGGAGCTGCCCGTACCAGCGTTACTGACGGCAGATCCGCGTGG
CATCATTAACATTGCCGTGCAGCAAGGCCAAACCCAGTGGATGGGCAAATCCGTGACTACGTGGGGATAACAAC
GGTAATCTCTTAGGCCCGGCAATTCAGCTGGATCGCGGCAAACCGGTGACGTTGAATTTGCACAATACCCTGC
CGGAAGCGACAACCATTCATTTGGCATGGCTTAGCGTTACCAGGCGAAGTAGATGGTGGTCCACAAGCTGTGAT
CGCCCCAGGAGCTTTTCGTCGGGTAAGTTTTACGCCGATCAGCCTGGTGTACATGCTGGTTTTCATCCGCAT
CAGCATGGTGCCTGTTACCAGGTTGCGCAAGGTCTGGCGGGACTGGTGTCTTAAAAGATGACGCAGGGC
AGAACTGCTGCTTCCGAAAATCTGGGGAGTTGACGATATCCCGTCATCTGCAGGATAAACGCCTGTCCAC
CGATGGCAGCAAAAATCGATTATGCGCTGGACATGATGTCAGCCACTGTCCGTTGGTTTGGCAACACCATGTTG
ACCAACGGCGTCATTTATCCACTCCAAGCTGTTCCACGTGGGTGGTTACGCCTCCGCTTGTCTAATGGGTGCA
ATGCTCGTAGCCTGAATCTCACCACATCGGATCAGCGCCCGCTGTATGTGATTGGTAGCGATGGTGGGTTGCT

GGCTGAGCCGGTAAAGCTGACCGAACTGTCCATGATGCCAGGGGAACGCTTTGAAGTTCTCGTTGACACCCAC
GATGGCAAAACCTTCGACCTTCAGACGCTGCCGTTTCGCCAAGCCGGAATGACGCTTGCCCCGTTTGATCAGC
CATTACCGCTGCTGACCATCCAGCCTTTGCAGATTTCTGGCCTCCGGCAGTCTGCCGGACAAACTGGTGGACAT
GCCAGCGTTTACCGGCTCACGACGGTGTCAAAGAACGTTGGCTTCAGTTGATGATGGACACGGAACCTGGATGAC
CGCGGTATGCAGGCCCTGATGGAGAAATGATGGCCGTAGTGCCTGGCGGGTAGTCGCCATGATGCGCACAACA
CAGATGGCGATATGTCGTCTGCAGTCGGTAAAGGCATGGACCACGATGCAATGGCCGGTATGGCCAATGGCTC
TATGCCGGGGATGGCCACAATCTGCCCGCGCAGAAACCGTACGATTTCCATCACGGAACAAGATCAACGGC
GTTGCGTTCGACATGAACACTCCGAGCTTTGACGTCAAACAAGGCGCACCTGGAGAAATGGACTATTTCTGGCG
AAGGGGATGGGATGTTGCATCCTTTCCACATTCATGGTGCAGTTCGCAATTCGAGCGAAAACGGTAAACC
CGCAGCTGCGCATCGGTGAGGATGGAAGGACATGGTGGTGTGGAAGGCTGGCGCTCGGAAGTGTGGTTTCGC
TTCAACCATCTTGCCACGAAAGATCAGGCGTATATGGCACATTGTCATCTCCTGGAACACGAGGATACCGGTA
TGATGCTCGGCTTTACCGTATCTGCGGAGCTC

Ips14138

Nucleotide sequence (1575 nt):

ATGCATCGTCGTGATTTTTCTGAAATATTTCTGCTGCTTTTTGGCGCATTTCAGCGCGCTGCCGCTCTGGAGCCGTA
CGGCATTGGCTGCGGAACGCCCTATTTTTGCCCTGTCCCTGAATTACTTGCACCCGATGCCAGAAACAGCATCCG
TCTTACGGCCAGGCCGGGAAAACCGTCTTCGGCGGCAAACTGCGACAACCTGGGGCTATAACGGCAACTTG
TTAGGGCCAGCGCTGCGCCTGACCCAGGGTGAAAGCGTTACCGTAGATATTCATAATAGCCTGGCTGAAGAAA
CAACGGTGCATGGCATGGTCTGGAAGTGCCGGGCGACGTCGACGGCGGCCCGCAGGGTGTGATTGCCGCCGG
GGGCAAGCGCACCGTCAAATTCACGCCTCAGCAGCGTGCCGAACCTGCTGGTTCCATCCCCATCAGCACGGT
AAAACCGGACATCAGGTGGCAATGGGGCTCGCGGGGCTGGTGTGATCGAAGACAGCGAAAGCCGACGGCTGA
TGCTGCCGAAGCAGTGGGGCATCGACGATATCCCGCTCATATTCAGGACAAGCGTTTTGGCGCCGACGGCGA
GATTGATTACAAGCTGGACGTGATGAGCGCCCGCTTGGCTGGTTTTGGCGACACGCTGCTGTGCAACGGTGT
AATTACCCGACGATTTCCAATCCTCGCGGTGGCTGCGTTTTACGCCTGCTTAACGGCTGTAACGCCCGTTTCAT
TGAATATTGCGGCAAGCGACAATCGCCCTCTGTATGTTATCGCCAGCGACGGCGGCCCTGCTGGCCGAACCGGT
GAAAGTCACGGAGCTGCCGCTGGTGTGAGCGCTTTGAAGTGTGGTGGACACCAGCGACGGTAAGCCG
TTTGATATCGTGACCTGCCGGTGAAGCAAATGGGTATGACCGTTGCGCCGTTGACCCAGCCGACGCCATCG
CGCGTATTCAGCCGGTGCATTTGCTGCTTCTGGCGAGCTGCCGGATAAGCTGGTGGAGGTTCCCCTCTGCC
AGCGCTCGAGGGCGTGACCGAACGCTGGTTACAGCTGATGATGGACCCCATGCTCGACATGATGGGCATGCAG
GCGCTGATGGAGAAATACGGCGAGAAAGCGATGGCGGGCATGAGCATGGCCGGCCACGGTAGTATGGCGGCA
TGAAGCAGGGCGGTATGGATCACGGCAAGATGAATCATGGTCAGATGGGCCAGGGTTTTGACTTCCACAATGC
CAACAAGATCAACGGCAAGGCCCTTTGATATGGCCACCCCGCGCTTCGCCCGCGAGAAGGGTAAATACGAGAAA
TGGACCATTTCCGGGGAAGGGGACATGATGCTGCATCCGTTCCACATCCACGGCACGCAGTTCCGCATTCTGA
GCGAGAACGGTCAGCCGGTTGCTGCCCATCGTCAGGGCTGGAAGATAACGGTTTCGTGTGGAAGGGGCGCGGAG
CGAGGTGTTAGTACGCTTCGATCACGAAGCCTCGAAAGAACATGCCATATATGGCCACTGTCATCTGCTGGAG
CATGAAGACACCGGCATGATGCTTGGCTTACCCTGGCATAA

Amino acid sequence (524 aa): 57.3 kDa (runs at ~62 kDa)

MHRRDFLKYSAAFSAFSAFLPLWSRTALAAERPI LPVPELLAPDARN S IRLTAQAGKTVFGGKTATTWGYNGNL
LGPALRLTQGESVTVDIHNSLAEETV **HWH**GLEVPDGDVGGPQGVIAAGGKRTVKFTVPPQRAATCWF **HPHQHG**
KTGHQVAMGLAGLVLI EDSESRRLMLPKQW GIDDIPLIIQDKRFGADGEIDYKLDVMSAAVGVFGD TLLCNGV
NYPQHSNPRGWLRLRL LNCNARSLNIAASDNRP LYV IASDGLLAEPVKVTELP LVMGERFEVLV DTS DGKP
FDIVTL PVKQMGMTVAPFDQPP IARIQPVRI AASGELPDKLVEVPALPALEGVTERWLQLMMDPMLDMMGMQ
ALMEKYGEKAMAGMSMAGHGSMGGMKQGGMDHGKMNHGQMGQGFDFHNANKINGKAFDMATPAFAAQKGYEK
WTISGEGDMM **HPFH**HTGTQFRILSENGQPVA AHRQGWKDTVRVEGARSEVLVRF DHEASKEHAYMA **HCHLLE**
HEDTGMMLGFTVA

Ips21622

Nucleotide sequence (1608 nt):

ATGCAACGACGCGATTTTTTGAATATAGCGCATTCTTAGGTGCTGCCACGATGCTGCCGGCATGGAGCCGGT
TTGCCTTTGCGGCGCAGGAACGTCCGGCACTCGGTATTCGCCCGCTGTAAACGGCGGATAATGCACAAAAAAT
CACCCCTGAATATTCAGCAGGGGCAAATGGTGTATTCGGGAAAAACCACCCAAACCTGGGGATATAACGGT
GACCTGCTGGGCCCCGGCGCTGAAACTGCGGCGTGGTAAGCCGGTAACGGTGGATATCGTCAATAAATGCGCG
AAGCGACAACCGTCCACTGGCACGGCCTGGAAATCAGTGGTGAACAGGATGGCGGTCCGCAGGCGATTATTGC
CCCGGGCAGCACCCGTACGGTCAGTTTCACGCCGGATCAGCCGGAAGCCACCTGCTGGTTCCATCCGCACACC
CACGGCAAAAACCGGCCATCAGGTGCGCAATGGGACTGGCCGGACTGGTGTGATTGAAGATGATAATACCGCAT
CAACCGGCTGCCGGGTGAGTGGGGAACGGATGATATCCCGTGGTCTTGCAGGATAAGCGCCTGACCGCAGA
CGGTA AAAATTGATTATCAGCTGGATGTGATGTCTGCCGACGTCCGGCTGGTTTGGCGACACCATGCTGACCAAC
GGCGCGGTTTACCCGACGATATGGCACC GCGCGCTGGCTGCGCTGCGTCAACGGCTGTAATGCC
GCAGTGTGCGGATGGCGGCCGGTGACGGGCGTCCGCTGTATGTGATCGGCAGTGACGGCGCCGCTGCTGGCCGA
ACCGTAAAAGTGCAGGAGCTGCCGATGCTGCCGGGAGAACGTTTTTGAAGTGTGGTGGATACATCAGACGGG
AAAGTCTTTGATATTGTCACCTTACCGGTACGTACGATGGGAATGACACTGGCACCGTTTTGATGAGGCGCTGC
CGGTGCTGTCCATCCGTCCGACACTCGATGCCGACGCGGCAAAC TGCCGGATGTGCTGGCGGCGATGCCGGC
TCTGATCGATACTCCCGCTGCCGGTACGTGAATTCGCTGTGATGATGGATATGCGCCTTGATATGCAGGGC
ATGATGATGCTGACCGATAAAATACGGCCCGCAGGCGATGGCCGGAATGGGCATGCACGGCAACATGGGCGGCG
GCCATATGAGCGGCGGCATATGGGCGGAATGAAAGGCATGAGCCGAGGTATGCACAGTGGCGGCGGTCCTG
CGGCTCCGGCGATGTTGATCTGATGAACGCCAACAGTATCAACGGTATTCGGTTCTCAATGACCGAAAGCGCA
TTTGATGTGAAGCAGGGACAGGCGGAGCGCTGGGTTATCTCCGGCGTCCGGTACATGATGCTGCACCCGTTCC
ATATCCACGGCACCCGTTTCCGTATTCTCAGTGAGGGCGGACGCACACCGGGCGGCTCACCGGCAGGGCTGGAA
AGATATGGTCAGGATAGAGGGCAATGTCAGTGAGGTTCTGGTGAATTTGACCATCCGGCCACTAATGCGCAT
CCGTTTATGGCGCACTGTCACCTGCTGGAACATGAAGATACCGGTATGATGACGGGCTTTACTGTCACGGCCT
GA

Amino acid sequence (535 aa): 58 kDa (runs at ~62 kDa)

MQRDFLKYSAFLGAATMLPAWSRFAFAAQERPALGIPPLLTADNAQKITLNIQQGMVFI PGKTTQTWGYNG
DLLGPALKLRRGKPVTVDIVNKLPEATTVHWHGLEISGEQDGGPQAI IAPGSTRTVSFTPDQPEATCWFHPHT
HGKTGHQVAMGLAGLVLIEDDNTASTGLPGEWGTDDIPVVLQDKRLTADGKIDYQLDVM SAAVGFWDMLTN
GAVYPQHMAPRGWLRRLRLLNGCNARSLRMAAGDRPLYVIGSDGGLLAE PVKVQELPMLPGERFEVLVDTSDG
KVFDIVTLPV RQMGMTLAPFDEALPVLSIRPTLDAARGKLPDVLAAMPALIDTSRLPVREFRLMMDMRLDMQG
MMLLTDKYGPQAMAGMGMHGNMGGGHMSGHMGMSRGMHSGGGHCGSGD VDLMNANSINGI PFSMTESA
FDVKQQAERWVISGVGDMMLHPFH IHGTRFRILSEGGRTPAHRQGWKDMVRIE GNVSEVLVKFDHPATNAH
PFMAHCHLLEHEDTGM MTGFVTVA

Ips24328

Nucleotide sequence (1605 nt):

ATGAACCGTCGTGATTTTCGTAAAAATGGACAACCCCTGATGGGGGCCCGCAGTACGCTGCCCGGCTGGAGCCGCT
TCGCGCTCGCCGCAGATCGCCCTGCATTACCCATTCTGCGCTGCTGGAACCGGATATCCGTAACGCCATTAT
GCTGACGTTGACAGCGGGTCAGAGCCAGTTTCTGCCGGGTGTAACACCGAAACCTGGGGCGTTAACGGTAAT
CTTCTCGGCCCGGCGCTGCGTATCCGTCGCGGCAAACAGGTCGATGTCACCGTCAATAACCGTCTGGATGTTG
CCAGCACCGTTCACTGGCACGGGCTGGAAATCCCGGTGACGTCGATGGCGGTCCCTCAGGCACTGATCGCTCC
AGGTCAGAAAAGAAAAGTCAGTTTCACTCGATCAGCCAGCGTCGACCTGCTGGTTCCATCCGCATCCGCAT
CAGACCAGCGTTATCAGGTGGCGATGGGGCTGGCCGGTATGGTGTGATTGAAGATGAAGCCAGCGACTTGT
TGCAGATCCCTAAACGTTGGGGCGTGGATGATATTCGGTCAATTTTGCAGGATAAACGCCTGAACGACGCCGG
ACAGATTGATTATCAGATGGACGTGATGACCGCGGCTGTCGGCTGGTTCCGGCAACATATGCTGACCAACGGC
GCGGTGTATCCGCAGCACGGTATTTCCCGCGGCTGGGTGCGTTTTTCGTCTGCTCAATGGCTGTAATGCACGCT
CGCTGCACATCGCCACCAGTGACCAGCGGCCGATGTACGTCATCGCCAGCGACGGCGGTTTCTGCTGAACC
GGTGA AAGTCAGCGATTTATCCTTGCTGATGGGCGAGCGTTTTTGAAGTGTGATTGATTGCAGCGACGGCAA
GCCTTTGATCTGGTTACGCTGCCGGTGAACAGATGGGTATGACGCTGGCACCTTTTCGACAAACCGTTGCCGG

TGCTGCGTATTTCAGCCGACGCTGACGCAAAGCGGCAGTTCTCTGCCGGACACGCTGGTCCCCTGCCGGCGCT
GGTGTCCGTCGATAATCTGCCAACCCGCTGGCTGCAACTGATGATGGATCCGCAACTGGATCAGCAGGGCATG
GCGGCGCTGATGAAACGCTATGGTCCACAAGGCAATGGCCGGCATGAGCATGGATCATGGCGGGCGGATATGG
CGGCAATGCCGGGCATGTCCGGTTCAGAACATGAGGGTACGGCAGCATGGCGGGTATGGATATGAGCAAATC
GTCCGCCGGTTACGACTTCATGCAGGGCAACAAAATTAACGGCAAAGCCTATGACATGAATGTACCCGCTTTC
GATGTGAAACAAGGCCAGTACGAAAAATGGACCATTTCGGCGGAAGGCGACATGATGCTGCATCCTTTCCATA
TCCACGGCACGCAGTTCGGTATTCTTTCTGAAAAATGGTCAGCCTGTGCCGCCGACCCGGCAGGGCTGGAAAGA
TATCGTGCGCGTTGAGGGTGCCCGCAGTGAAGTCTGGTGCGCTTCAACCATCTGGCCAATAAACAACATGCT
TATATGGCGCACTGTCATCTGCTGGAACATGAAGATACCGGCATGATGCTTGGATTTACGGTATCGGCCGTAA

Amino acid sequence (534 aa): 58.8 kDa (runs at ~63 kDa)

MNRRDFVKWTTLMGAASTLPGWSRFALAADR PALPI PALLEPDIRNAIMLTLQRGQSQFLPGVNTETWGVNGN
LLGPALRIRRGKQVDVTVNRLDVASTVHWHGLEIPGDVDGGPQALIAPGQKRKVSFTLDQPASTCWFHPPH
QTSGYQVAMGLAGMVLIEDEASDLLQIPKRWGVDDIPVILQDKRLNDAGQIDYQMDVMTAAVGFQGHMLTNG
AVYYPQHGISRGWVRFRLNLCNARSLHIATSDQRPMYVIASDGGFLPEPVKVSDDLMLMGERFEVLI DCSDGK
AFDLVTLFVKQMGMTLAPFDKPLPVLRIQPTLTQSGSSLPDTLVPLPALVSDNLPTRWLQLMMDPQLDQQGM
AALMKRYGHSAMAGMSMDHGGGDMAAMP GMSGSEHEGHGSMAGMDMSKSSAGYDFMQGNKINGKAYDMNVPAF
DVKQGGYQEKWTISGEGDMLHPFHHTGTQFRILSENGQVPPHRQGWKDIVRVEGARSEVLVRFNHLANKQHA
YMAHCHLLEHEDTCMMLGFTVSA

Ips28714

Nucleotide sequence (1611 nt):

ATGAACCGTCGCGATTTCTGTAAGTGGACAACGCTCCTGGGCGCCGCCAGTACGTTACCCGGCTGGAGCCGGT
TTGCGCTGGCCGCTGACCGCCCGGCTTGCCATTTCCACACTGCTTGAACCGGATGCCCGCAGCGCGATTTTC
GCTTTCCCTGCAACGTGGTCCAGAGCCAGTTTCTGCCCGGGTGAATACACAAACCTGGGGCGTGAACGGCAAT
CTGCTGGGGCCTGCGCTGCGCGTGCGGGCGGGGCAAAGCGCTCAATGTCACGGTCAAAAACAATCTGGACGTCG
CGAGCACCGTTCACTGGCACGGGCTGGAAAATTCCTGGTGACATTGACGGCGGTCTCAGGGGCTGATCGCGCC
GGGCCAGAGCCGACGCGTGAAGTTGTCAGCTCGATCAACCCGCCGCGACCTGCTGGTTTTACCCGCATCCGCAT
CAGACCAGCGGCTATCAGGTGGCAATGGGGCTGGCCGGATTAGTCTGATTGAAGATGAAGGCAGCGACGCAT
TACAGATCCCAGAACGCTGGGGCGTGGATGATATCCGGTCAATTTGTCAGGACAAACGCCTCAGCGACGCCGG
ACAGATTGATTACCAGATGGACGTGATGACTGCCCGCATCGGCTGGTTCGGCGAGCACATGCTGACTAATGGC
CGGATTTATCCGACACCGGCATTTTCGCGCGGCTGGTTCGTTTTCTGCTGCTTAACGGCTGCAATGCCGCT
CGTGCACATCGCCACCAGCAGCGCCGATGTACGTTATTGCCAGCGACGGCGTTTTCTTTCCGAGCC
GGTAAAAGTCAGTGATTTGCCGTTGCTGATGGGGGAACGTTTTTGAAGTGTGATTGATTGCAGCGACGGTAAA
GCGTTTTGATCTGGTACATTTGCCGGTGAAACAAATGGGTATGACGCTGGCACCGTTTCGACAAGCCGTTGCCGG
TACTACGCATTCAGCCGACATTTGACCCAAAAGCGCGGTTCACTGCCGGACACGCTGGTGCCGCTTCCGGCGCT
GGTTTTACTGATAATCTGCCGACCCGCTGGTTGCAACTGATGATGGATCCGCAACTGGATCAGCAGGGCATG
GCGGCGCTGATGAAACGTTACGGTACAGCGCGATGGCCGGCATGAGCATGGATCACGGCGGGCGGTGATATGG
ATATGGCGCGGATGCCGGGGATGTCGGTTCGGGCCATGACGGCCACGGCGGTATGGCCGGAATGGACATGGG
CAAAGTGGCCGGAAGTTACGATTTTCATGCAGGGCAATAAAATCAACGGCAAAGCCTACGACATGAACGTGCCT
GCGTTTCGATGTGAAACAAGGGCAATACGAGAAATGGACGATTTCCGGCGAAGGCGACATGATGCTGCATCCGT
TCCATATCCACGGCACACAGTTTCGTATTTCTTTCTGAAAACGGCCAGCCGGTTCGCGCCGACCCGTGAGGGCTG
GAAAGATATCGTCCGCGTCGAAGGCGCCCGCAGCGAAGTGTGGTGGCTTTCAGTACCTGGCGGATAAACAG
CATGCCTATATGGCGCATTTGCCATCTGCTGGAACACGAAGATACCGGCATGATGCTCGCGTTTTACGGTGTCCG
CATAA

Amino acid sequence (536 aa): 58.4 kDa

MNRRDFVKWTTLLGAASTLPGWSRFALAADR PALPIPTLLEPDARSAISLSLQRGQSQFLPGVNTQTWGVNGN
LLGPALRVRRGKALNVTVKNLNDVASTVHWHGLEIPGDIDGGPQGLIAPGQSRVSLQLDQPAATCWFHPPH
QTSGYQVAMGLAGLVLIEDEGSDALQIPKRWGVDDIPVILQDKRLSDAGQIDYQMDVMTAAIGWFGEHMLTNG
AIYPQHGISRGWVRFRLNLCNARSLHIATSDQRPMYVIASDGGFLSEPVKVSDDLPLLMGERFEVLI DCSDGK
AFDLVTLFVKQMGMTLAPFDKPLPVLRIQPTLTQSGGSLPDTLVPLPALVSTDNLPTRWLQLMMDPQLDQQGM
AALMKRYGHSAMAGMSMDHGGGDMDMAAMP GMSGSGHDGHGGMAGMDMGKVAGSYDFMQGNKINGKAYDMNV

AFDVKQGQYEKWTISGEGDMMMLHPFH IHGTQFRILSENGQPVAPHRQGWKDIVRVEGARSEVLVRFSHLADKQ
HAYMAHCHLLEHEDTGMMLAFTVSA

Optimized nucleotide sequence for expression in *E. coli*:

GCTAGCATGAATCGCCGTGACTTCGTTAAATGGACGACCTTACTCGGGGCTGCGTCGACCTTACCAGGCTGGT
CACGTTTTGCGCTCGCTGCGGATCGGCCGGCATTGCCCATTCCTCAACTCTGTTGGAACCTGATGCACGTAGCGC
CATTTCCTCAGTCTGCAGCGCGGTCAAAGCCAGTTTCTGCCGGGTGTTAACACACAGACCTGGGGCGTGAAT
GGCAACCTCCTGGGACCGCATTACGGGTACGTCGTGGCAAAGCGCTTAACGTCACCGTGAAGAACAATCTGG
ATGTTGCGTCAACCGTGCACCTGGCACGGCTGAAAATTCCGGGTGATATTGATGGCGGGCCCCAAGGCCTGAT
CGCTCCAGGTCAGTCTCGCTCGGTAAGTCTGCAACTGGATCAACCGGCTGCGACCTGTTGGTTTTACCCACAT
CCACATCAGACCTCTGGCTATCAGGTGGCGATGGGCCCTTGCCGGTCTGGTGTGCTGATCGAGGATGAAGGGAGCG
ATGCGTTGCAGATTCCGAAACGCTGGGGCGTCGACGACATCCCCGTTATTCTTCAGGATAAACGCCTGTCTGA
TGCCGGTCAGATTGACTACCAGATGGATGTCTGACTGCAGCTATTGGATGGTTCGGTGAGCACATGCTCACC
AATGGCGCCATCTATCCGCAACACGGCATTTCACGTGGCTGGGTTCGCTTTCGCTTACTGAACGGTTGTAATG
CCCGCAGCTTACACATCGCGACATCGGACCAACGCCCGATGTATGTGATTGCGAGCGATGGTGGATTCCCTGTC
TGAACCTGTCAAAGTCTCCGATCTGCCTCTGCTGATGGGCGAACGCTTTGAAGTGTGATCGATTGCTCGGAC
GGCAAAGCCTTCGATCTGGTCACACTTCCGGTGAACAGATGGGGATGACGCTGGCGCCGTTTCGATAAACCCAC
TGCCGGTATTACGATTCAGCCGACCTGACTCAGAGTGGTGGTTCCCTGCCGTGATACGCTGGTTCCTCCCTGCC
GGCCCTGGTGTCAACCGACAATTTGCCAACCCGTTGGCTCCAGCTGATGATGGACCCGAGTTGGATCAGCAG
GGTATGGCCGCACTTATGAAACGCTATGGACACAGTGCATGGCCGGTATGAGCATGGACCATGGTGGAGGTG
ATATGGACATGGCGGCGATGCCGGGCATGTCTGGCAGTGGCCATGATGGACATGGTGGCATGGCGGGTATGGA
TATGGGGAAAAGTGGCAGGCTCCTACGACTTTATGCAGGGGAACAAGATCAACGGGAAAGCCTACGACATGAAC
GTCCCGGCTTTCGACGTGAAGCAGGGCCAATACGAGAAATGGACGATTTCCGGTGAAGGGGATATGATGCTGC
ATCCGTTTCACATCCATGGTACTCAGTTTCGCATCTTGTCCGAAAATGGCCAACCGGTTGCACCTCATCGTCA
AGGCTGGAAGGACATTGTACGTGTTGAAGGTGCTCGCAGCGAGGTACTGGTTCGTTTTAGCCATCTGGCCGAT
AAACAACATGCCTATATGGCGCATTGCCACTTACTTGAGCATGAAGATACGGGGATGATGTTGGCATTACCG
TGAGCGCCGAGCTC

Declaration on oath / Eidesstattliche Versicherung

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ort/Datum – City/Date

Unterschrift – Signature

Acknowledgements / Danksagung

Zunächst möchte ich Prof. Dr. Wolfgang Streit herzlich danken, für die Chance meine Promotion in seiner Arbeitsgruppe durchzuführen, die Vergabe dieses spannenden Forschungsthemas und die freundliche Betreuung – auch über meine Anstellung an der Universität Hamburg hinaus. Außerdem danke ich Prof. Dr. Stefan Linder für die augenblickliche Bereitschaft, meine Arbeit zu begutachten.

Dr. Christel Vollstedt und allen Mitarbeitern der Abteilung für Mikrobiologie und Biotechnologie bin ich sehr dankbar für die Hilfe und Unterstützung während meiner Zeit in Klein Flottbek.

Martin Reuter danke ich sehr für die Aufstellung der Pheromonfallen und Sammlung der Borkenkäfer, die eine unverzichtbare Basis meiner Arbeit darstellen.

Ein großes Dankeschön geht auch an Ulf Tabel, Beate Kogge und Angela Wichmann für die Motivation und dafür, dass sie mir immer den Rücken freigehalten haben, wenn ich Zeit zum Schreiben brauchte.

Kleine und klitzekleine Maid, ohne euch wäre ich oft verzweifelt und die Zeit im Labor wäre nicht so schön gewesen. Vielen Dank für immer offene Ohren, die Aufmunterungen und den Beistand. Ich bin so froh, dass wir das alles zusammen durchgestanden haben!

Letztlich danke ich meiner Familie und Matthias unendlich für ihre Geduld, Nachsicht und Unterstützung – und besonders meinem Papa, der die Fertigstellung meiner Arbeit leider nicht mehr miterleben konnte.