



Article

# Valorization of Biomasses from Energy Crops for the Discovery of Novel Thermophilic Glycoside Hydrolases through Metagenomic Analysis

Roberta Iacono <sup>1,\*</sup>, Andrea Strazzulli <sup>1,2,†</sup>, Rosa Giglio <sup>3</sup>, Federica Bitetti <sup>3</sup>, Beatrice Cobucci-Ponzano <sup>3</sup> and Marco Moracci <sup>1,2,3,\*</sup>

<sup>1</sup> Department of Biology, University of Naples “Federico II”, Complesso Universitario di Monte S. Angelo, Via Cupa Nuova Cinthia 21, 80126 Naples, Italy

<sup>2</sup> Task Force on Microbiome Studies, University of Naples “Federico II”, 80138 Naples, Italy

<sup>3</sup> Institute of Biosciences and BioResources—National Research Council of Italy, Via P. Castellino 111, 80131 Naples, Italy

\* Correspondence: roberta.iacono@unina.it (R.I.); marco.moracci@unina.it (M.M.)

† These authors contributed equally to the work.

**Abstract:** The increasing interest for environmentally friendly technologies is driving the transition from fossil-based economy to bioeconomy. A key enabler for circular bioeconomy is to valorize renewable biomasses as feedstock to extract high value-added chemicals. Within this transition the discovery and the use of robust biocatalysts to replace toxic chemical catalysts play a significant role as technology drivers. To meet both the demands, we performed microbial enrichments on two energy crops, used as low-cost feed for extremophilic consortia. A culture-dependent approach coupled to metagenomic analysis led to the discovery of more than 300 glycoside hydrolases and to characterize a new  $\alpha$ -glucosidase from an unknown hyperthermophilic archaeon. Aglu1 demonstrated to be the most active archaeal GH31 on 4Np- $\alpha$ -Glc and it showed unexpected specificity vs. kojibiose, revealing to be a promising candidate for biotechnological applications such as the liquefaction/saccharification of starch.

**Keywords:** CAZymes; extremozymes; archaea; enrichments



**Citation:** Iacono, R.; Strazzulli, A.; Giglio, R.; Bitetti, F.; Cobucci-Ponzano, B.; Moracci, M. Valorization of Biomasses from Energy Crops for the Discovery of Novel Thermophilic Glycoside Hydrolases through Metagenomic Analysis. *Int. J. Mol. Sci.* **2022**, *23*, 10505. <https://doi.org/10.3390/ijms231810505>

Academic Editor: Qunxin She

Received: 9 August 2022

Accepted: 8 September 2022

Published: 10 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Extremozymes from thermophilic microorganisms possess notable properties, such as thermostability, and robustness in catalytic activity, making them interesting for a plethora of commercial applications such as the chemical industry, bioremediation and biorefinery [1–4].

Global annual lignocellulosic generates nearly 731 and 709.2 million tons of wastes of rice straw [5] and wheat straw [6], respectively; and 5 and 200 billion tons of crop residue waste [7] and plant biomass [5], respectively. Lignocellulosic biomasses are rich in sugars, lipids, proteins, and vitamins and have a chemical composition that can support microbiological growth [8]. Indeed, lignocellulosic biomasses from agricultural crops or wood have been utilized as substrates in fermentation processes (e.g., as part of culture media) to yield commercially relevant compounds such as ethanol [9,10], organic acids [11,12], enzymes [13,14], polymers [15], hydrogen [16], especially as alternatives of starch producing crops, thereby avoiding competition with food production [17,18].

These feedstocks can have different origins, depending on the biomass availability in different geographical regions in order to cut the costs and CO<sub>2</sub> emissions due to their transportation in the biorefineries plants. In a study performed in Italy, it has been demonstrated that the plant and the transportation of biomass are the main contributors to the production of CO<sub>2</sub>, accounting for about the 38% and 34% of the total, respectively [19].

*Arundo donax* and *Cynara cardunculus*, two native Mediterranean plant species, have been recently identified as potential energy crops in this area. *C. cardunculus*, also known as wild cardoon or Castilian thistle, is a perennial herbaceous species with an annual growth cycle. The branched stalks of cardoon account for about 40% of total dry biomass and are composed generally of 20–30% of hemicelluloses (made up of arabinoglucuronoxylan), 35–45% of cellulose and 10–20% of lignin [20]. In addition, in *C. cardunculus* inulin is often stored in specialized organs, such as taproot, bulbs, and capitula and used as a carbon source during regrowth and sprouting in the spring [21]. *A. donax*, also known as giant reeds, is a perennial rhizomatous wetland grass with high content of cellulose, hemicellulose (composed of arabinoglucuronoxylan), and lignin, by 31%, 35% and 18%, respectively [22]. In addition, the most abundant non-structural carbohydrates in this energy crop are sucrose and starch [23].

Due to the different types of polysaccharides present in these herbaceous species and to their structural complexities, complete enzymatic deconstruction requires the synergistic action of several carbohydrate active enzymes (CAZymes). “Omics” approaches have provided a powerful tool for the discovery of new extremozymes from nature [24]. In fact, most extremophilic microorganisms are recalcitrant to lab cultivation and isolation approaches [4], therefore, culture-independent metagenomic strategies are promising approaches to assess the phylogenetic composition and functional potential of microbial communities living in extreme environments [25]. Recently, our group reported on the metagenomic analysis of the microbial communities populating the Pisciarelli hot springs (Naples, Italy, 40°49′45.1″ N; 14°8′49.4″ E) [26,27], identifying the repertoire of carbohydrate active enzymes (CAZome) produced by the microbial community of extremophiles populating this environment [27]. Pisciarelli’s microbiome showed a huge number of genes encoding putative CAZymes, which include glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and auxiliary activities (AAs) [27], which are classified in the CAZy database ([www.cazy.org](http://www.cazy.org); accessed on 9 August 2022) [28].

The main aim of this study was to select new CAZymes by enriching Pisciarelli’s samples on *A. donax* and *C. cardunculus* as low-cost rich carbon sources as an alternative to the common culture media, and as possible inducers of growth of specialized microorganisms. Since Pisciarelli solfatara is mainly populated by Archaea strains that are often difficult to isolate in laboratory conditions, we follow a metagenomic approach to analyze the efficacy of the enrichments and to identify the enzymes of interest. We report here that this approach led to the discovery of hundreds of new Glycoside Hydrolases (GHs) and the characterization of a thermophilic and thermostable  $\alpha$ -glucosidase from an unknown hyperthermophilic archaeon.

## 2. Results

### 2.1. Enrichments

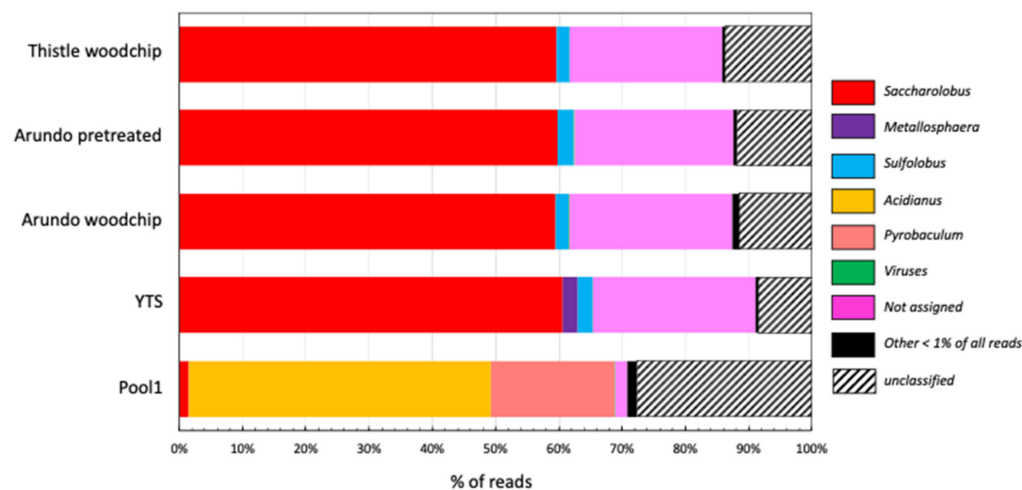
The mud/water sample collected in Pisciarelli solfatara (Pool1, pH 5.5 and 85 °C) in March 2012 [27] was incubated in Basal Salt Medium (BSM) at pH 1.8 [29], Brock salt medium pH 3.5 [30], or Pyrobaculum (PYR) salt medium pH 7.0 [31] supplemented by yeast extract, tryptone and sucrose (YTS) 0.1% as an initial carbon source. *A. donax* (giant reed) pretreated by steam-explosion (0.15% *w/v*) was used in the first trial as alternative rich carbon source. Only the microbial enrichment in Brock medium with pretreated *A. donax* or with YTS 0.1%, showed an increase in optical density, while no microbial growth was observed in BSM and PYR salt media containing the same sources of energy (Figure S1).

After this preliminary trial, a fresh culture was set up by starting from the same environmental sample in Brock salt medium supplemented by YTS 0.1% as an initial culture broth. After 24 h, the culture was split into four sub-cultivations and YTS was substituted with *A. donax* or *C. cardunculus* (Thistle), both as pretreated and woodchip. Contextually, a further culture was performed by using YTS 0.1% as a carbon source. Growth was observed in all enriched cultures, except on pretreated Thistle. When the

enrichments reached 0.4 OD<sub>600</sub>, total DNA was purified from the cultures as described in Materials and Methods, and fully sequenced.

## 2.2. Sequence Analysis and Microbial Composition

The microbial composition of all the enriched samples (Arundo woodchip, Arundo pretreated, Thistle woodchip, and YTS), was analyzed by performing a metagenomic sequence-based approach by Illumina (San Diego, CA, USA) HiSeq sequencing (Table S1), followed by community diversity analysis based on assigning reads to known microorganisms (Figure 1).



**Figure 1.** Taxonomic annotation at the genus level.

Interestingly, in all samples, among the known genus, the most abundant is *Saccharolobus* (~60%) followed by *Sulfolobus* ranging from 2.5% in the sample containing Arundo pretreated to 2.0% in Thistle. More specifically, the reads assigned to the genus *Saccharolobus* are related to the species *S. solfataricus* (57%), *S. shibatae* (0.6%), and *S. caldissimus* (0.1%). Differently, the genus *Metallosphaera* is present only in the YTS sample (2.4%) and, possibly it has been negatively selected by the other biomasses.

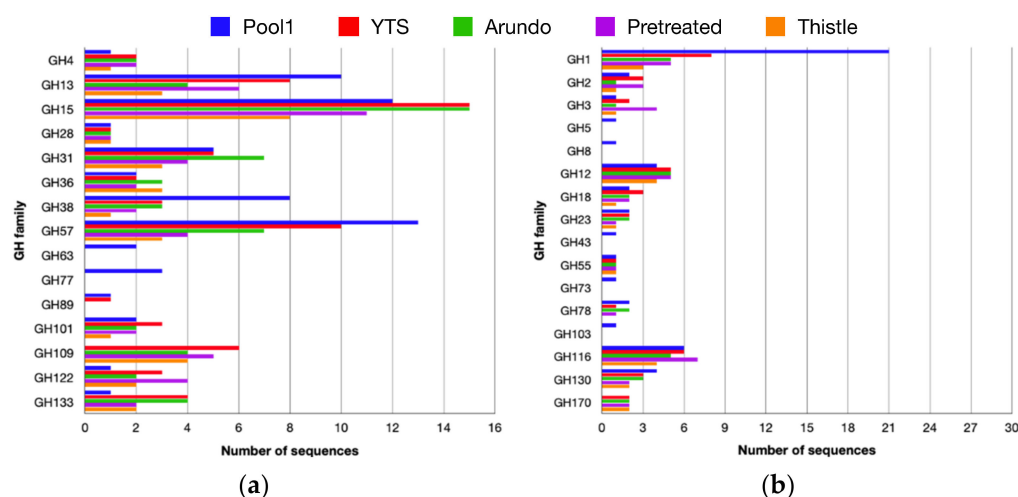
The carbon sources in the enriched samples drove a dramatic selection of the microbial community present in the original environmental sample of Pool1 (Figure 1). In fact, while Pool1 was dominated by the genera *Acidianus* (48%) and *Pyrobaculum* (19%), in the enriched samples the first one is present at 0.2% while the latter is completely absent.

It is worth noting that in all the enriched samples at the genus level, the percentage of not assigned reads and sequences that can be assigned to ranks higher than genera is ~25%, while at the ranks of phylum, class, order, and family, the percentage of unassigned reads is <0.1%. This indicates the presence of microorganisms belonging to the *Sulfolobaceae* family but different from the genera present in the Refseq NCBI database (Figure S2). In addition, in all the enriched samples, there are reads that could not be taxonomically classified (termed unclassified in Figure 1) representing 8.5%, 11.5%, 12%, and 13.8% of the reads of YTS, Arundo woodchip, Arundo pretreated, and Thistle woodchip, respectively, while in Pool1 these represented 28%.

The reads of each sample were assembled individually and 1338 contigs ( $\geq 500$  bp) were obtained from YTS, 1751 from Arundo woodchip, 1368 from Arundo pretreated, and 549 from Thistle woodchip. Subsequently, the analysis of Open Reading Frames (ORFs) allowed the identification of 9947 in YTS, 8334 in Arundo woodchip, 6439 in Arundo pretreated, 4300 in Thistle woodchip, and 14,934 in Pool1 encoding sequences for hypothetical proteins (Table S1).

### 2.3. CAZome Analysis, Cloning, Expression, and Purification of *Aglu1*

In order to evaluate if the enrichment on the different media modified the number and/or type of CAZymes, the previously identified ORFs were analyzed using the dbCAN2 pipeline [32]. The analysis revealed that among CAZymes, Glycosyl Transferases (GTs) (Figure S3a) represent the most abundant class (52% of the total number of CAZymes) grouped in 12 families, followed by GHs (39%) (Figure 2). In contrast, Carbohydrate Esterases (CEs) and Auxiliary Activities (AAs) (Figure S3b) are much less abundant (5% and 4%, respectively).



**Figure 2.** Distribution of the glycosidases active on axial (a) and equatorial (b) C1-OR bonds.

Among GHs, 23 families were identified in all samples, 14 of which were active on substrates with axial C1-OR bond (Figure 2a) and 11 on substrates showing equatorial C1-OR bond (Figure 2b) for a total of >300 sequences encoding for putative GHs. In general, selection produced a clear change in CAZymes composition in the four enrichments if compared to Pool1, as both number and type of families, and numerosity of enzymes. In particular, among the enzymes active on substrates with equatorial O-glycosidic bonds (C1-OR), different sequences belonging to families of exo- and endo-glycosidases active on hemicellulosic oligosaccharides such as xylan, xyloglucan, and mannan (GH1, GH3, GH12, and GH116) have been identified. Among the families grouping enzymes active on axial C1-OR bonds, 6 group exo- and endo-glycosidases were active on starch and maltodextrines (GH13, GH15, GH31, GH57, GH122, and GH133). It is worth noting that most of the GHs families active on substrates containing equatorial C1-OR bonds that have been found in the microbiomes growing on the four carbon sources tested were present in the original sample from Pool1. The number of sequences present in the GHs families active on equatorial C1-OR bonds from the enrichments was about the same as those found in Pool1. However, a clear selection was observed. Sequences from GH5, GH8, GH73, and GH103, which were present in Pool1, were absent in the enrichments and the number of sequences assigned as GH1 originally present in Pool1 was dramatically reduced upon enrichment. On the other hand, the sequences annotated in GH170 that could be identified in the microbiomes from all four enrichments were lacking in Pool1 and family GH3 showed 4 CAZyme sequences in Arundo pretreated vs. 1 in Pool1.

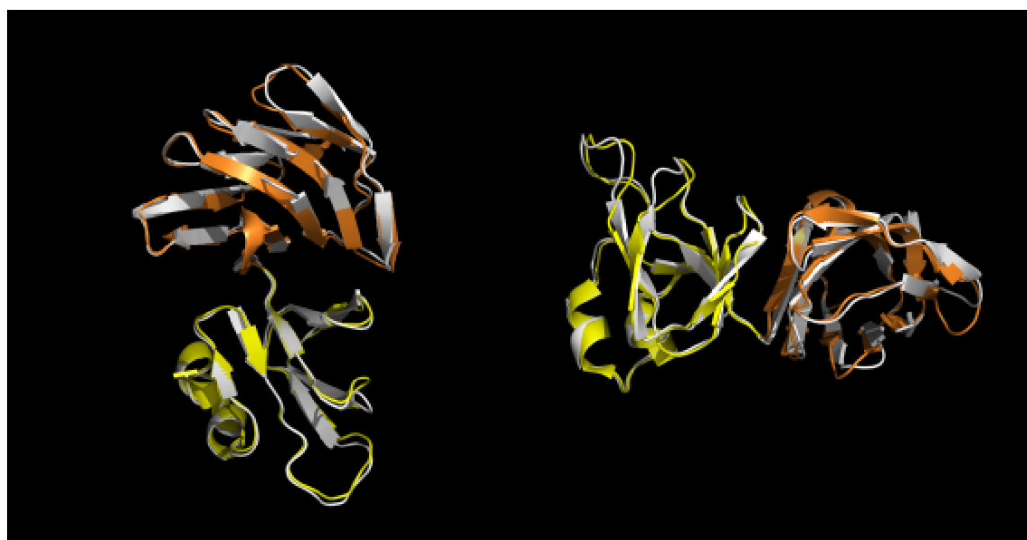
Among the selected GHs families active on axial C1-OR bonds, a higher variety was observed. Sequences in families GH122 and GH133 significantly increased in number if compared to those present in Pool1. GH109, which is absent in Pool1, showed several sequences in all four enrichments. In addition, sequences from two families, GH15 and GH31, grouping enzymatic activities involved in starch degradation, were more abundant in Arundo woodchip than in Pool1.

These observations suggested that enrichments produce a more significant selection on the CAZymes involved in starch degradation, rather than in lignocellulose hydrolysis.

This prompted us to analyze in more depth the sequences encoding for starch-degrading enzymes. In particular, a specific gene of 2082 bp present in all samples, named *aglu1*, and encoding for a hypothetical protein (Aglu1) of 693 aa belonging to family GH31, was identified in a contig of 21,649 bp (Figure S4b).

The multialignment of Aglu1 against other characterized archaeal representatives of the GH31 family, including MalA and XylS from *S. solfataricus* [33,34], ST2525 from *S. tokodaii* [35], MalA from *S. acidocaldarius* [36], AglA from *P. torridus* [37], and Agla from *T. acidophilum* [38] confirmed the presence of several conserved regions. In particular, the highly conserved catalytic residues of GH31 family members strongly suggested that *aglu1* could encode for a functional enzyme (Figure S5).

To date, the crystallographic study of the *S. solfataricus*  $\alpha$ -glucosidase MalA is notable as the first and sole experimental report of the 3-D structure of an archaeal (hyper)thermophilic GH31 [39], sharing 82% overall amino acid identity with Aglu1. In order to investigate the possible three-dimensional structure of Aglu1, a model was prepared using ColabFold [40] and compared to the MalA structure (PDB ID: 2G3M). As reported by Ernst and collaborators, MalA presents 4 domains (N, A, C, and D) and two subdomains (B and B') [39] (Figure S6a). The Aglu1 3D model conserved all MalA domains and subdomains with an overall RMSD of 0.461Å (Figure S6b,c). However, detailed analysis showed specific differences in the sequence identity of the five domains. In particular, the N-terminal (N), the catalytic domain (A), and the subdomains B and B', showed identities with MalA and other  $\alpha$ -gluco and xylosidases of the GH31 family in the range of 86–90% [39]. Instead, the C and D domains of MalA and Aglu1 had significantly lower sequence identity (68% and 50%, respectively) (Figure 3). In particular, it has been suggested that the D domain, highly heterogeneous in the family, could be involved in the binding of starch substrate [41].



**Figure 3.** Close-up superimposition of *S. solfataricus* MalA C (yellow) and D (orange) domains and Aglu1 model (white).

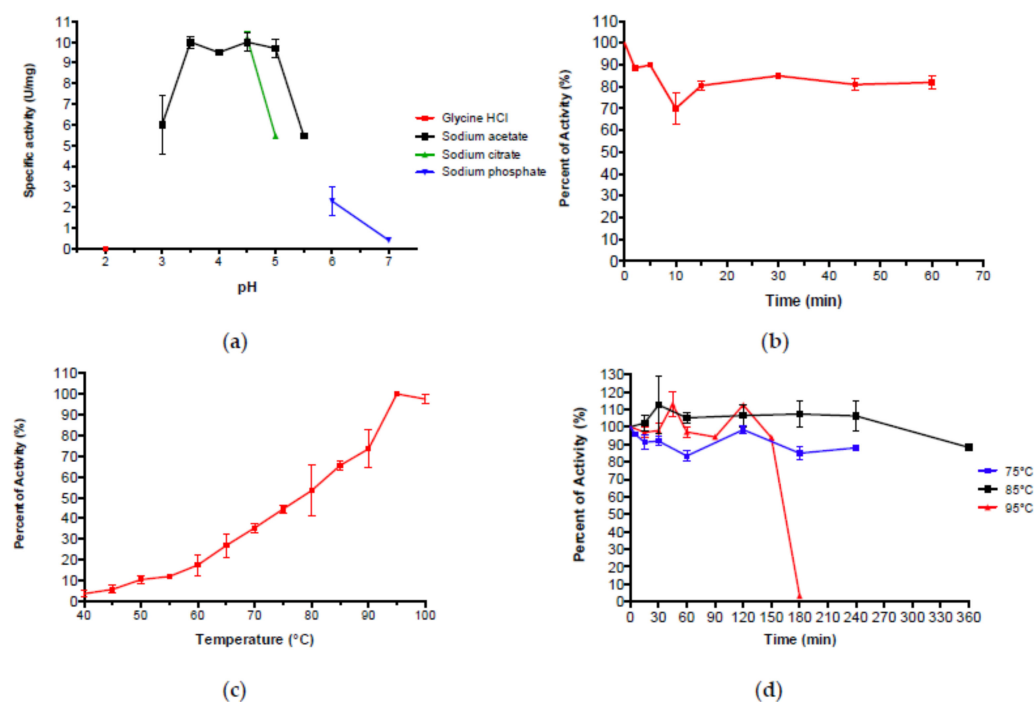
In order to understand if the observed structural diversity had functional implications, the gene *aglu1* was cloned in the pET20b expression vector and recombinantly expressed in *E. coli* BL21 (DE3) RIL. The recombinant Aglu1 was subsequently purified through six steps of purification, as described in materials and methods, with a final yield of 34% (1.25 mg/L of culture) and a purity of 95% (Table S2).

#### 2.4. Molecular Mass Determination, pH and Temperature Dependence

SDS-PAGE analysis of the recombinant purified Aglu1 revealed the presence of a band at the predicted molecular weight of 80 kDa (Figure S7). In native conditions, Aglu1 showed

a molecular mass of  $462 \pm 1.2$  kDa, as demonstrated by size exclusion chromatography indicating a hexameric structure (Figure S8) as observed for MalA [39].

Aglu1 standard assays were performed on pNp- $\alpha$ -Glc. The enzyme was optimally active in a sodium acetate buffer in the range of pH 3.5–5.0 (Figure 4a), as other thermophilic GHs showing a preference for slightly acidic pH [1,27,34,35,42]. PRED-TAT analysis indicated the absence of a signal peptide excluding that Aglu1 is a secreted enzyme. At pH 4.5, the enzyme retained 80% of its maximal activity up to 1 h at 65 °C (Figure 4b).



**Figure 4.** (a) Dependence of Aglu1 activity on pH. (b) Stability of the Aglu1 activity at pH 4.5. (c) Dependence of Aglu1 activity on temperature. (d) Heat stability of Aglu1.

As expected for an enzyme isolated from an enrichment performed at 80 °C, Aglu1 showed its maximal activity (~20 U/mg) at 95 °C which was reduced to 26%, and 66% at 65 °C and 85 °C, respectively (Figure 4c), as observed in archaeal  $\alpha$ -glucosidases from GH31 [33,34,36–38]. In addition, Aglu1 exhibited remarkable stability at 75 °C, 85 °C, and 95 °C maintaining 100% of activity after 240, 360 and 120 min, respectively (Figure 4d). These results demonstrated that Aglu1 is a highly thermostable acid  $\alpha$ -glucosidase.

### 2.5. Substrate Specificity and Kinetics Parameters

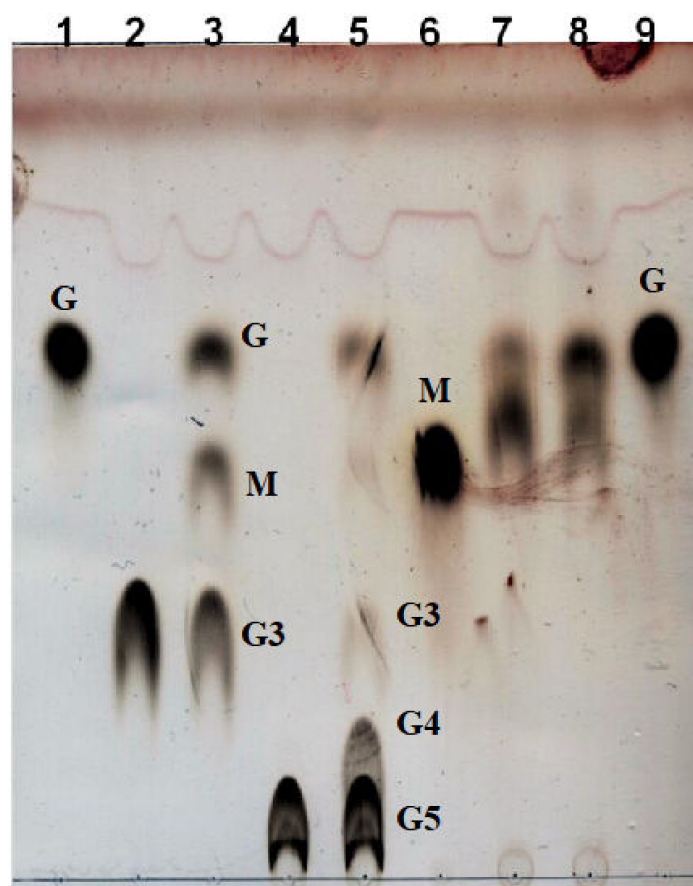
Preliminary substrate screening on aryl-glycosides, malto-oligodextrins, trehalose, nigerose and polysaccharides (starch and glycogen) at 65 °C for 1 h revealed that Aglu1 was able to hydrolyze  $\alpha$ -1,4-,  $\alpha$ -1,3- and  $\alpha$ -1,2- glycosidic bonds with a decreasing specific activity on higher maltodextrin DP, starch, and glycogen. The enzyme was not active on trehalose ( $\alpha$ -1,1-) and on the aryl-glycosides 4Np-  $\alpha$ -Xyl, - $\alpha$ -Fuc, - $\alpha$ -Gal, - $\alpha$ -Man, - $\alpha$ -L-Ara. Standard activity at 85 °C was measured on the best substrates 4Np- $\alpha$ -Glc, maltose, maltotriose and kojibiose showing  $22.2 \pm 0.6$  U/mg,  $73.9 \pm 1.3$  U/mg,  $73.3 \pm 5.0$  U/mg and  $54.0 \pm 3.0$  U/mg, respectively.

To investigate the substrate specificity, kinetic parameters were determined on the best substrates at 85 °C. Aglu1 showed the highest specificity constant on maltotriose followed by kojibiose, maltose and 4Np- $\alpha$ -Glu (Table 1 and Figure S9).

**Table 1.** Steady-state kinetic constants of Aglu1.

Substrates	$V_{MAX}$ (U mg <sup>-1</sup> )	$K_M$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )
4Np- $\alpha$ -Glc	42.5 $\pm$ 1.1	3.59 $\pm$ 0.33	327.4 $\pm$ 8.4	91.14
Maltose	117.1 $\pm$ 6.9	3.16 $\pm$ 0.65	901.6 $\pm$ 53.6	284.97
Maltotriose	89.7 $\pm$ 3.2	1.03 $\pm$ 0.20	690.7 $\pm$ 24.4	670.64
Kojibiose	65.0 $\pm$ 1.9	1.41 $\pm$ 0.19	500.2 $\pm$ 14.6	352.49

The ability of Aglu1 to hydrolyze maltotriose (G3), maltopentaose (G5) and nigerose was further investigated by thin-layer chromatography (TLC). As shown in Figure 5 (lanes 3 and 5), the enzyme produced glucose and maltooligodextrins shortened by one unit compared to the substrate, revealing that the enzyme is exo-acting. In addition, Aglu1 showed weak activity on nigerose (Glc- $\alpha$ 1,3-Glc), detecting a partial hydrolysis of the substrate after 1 h (Figure 5, lane 8).



**Figure 5.** TLC analysis of the enzymatic assays on maltotriose, maltopentaose and nigerose. Lanes 1 and 9: Glucose. Lane 2: blank on maltotriose. Lane 3: Aglu1 on maltotriose. Lane 4: blank on maltopentaose. Lane 5: Aglu1 on maltopentaose. Lane 6: Maltose. Lane 7: blank on nigerose. Lane 8: Aglu1 on nigerose. G: glucose; M: maltose; G3: maltotriose; G4: maltotetraose; G5: maltopentaose.

### 3. Discussion

The study aims to take advantage of the metabolic potentiality of the microbial communities inhabiting the site Pool1 in Pisciarelli solfatara hot springs by enriching the microbial communities on biomasses of biotechnological interest and analyzing the CAZymes that could be selected. A recent metagenomic study on the microbial population of Pool1 showed a remarkably high set of CAZymes that captured our attention for their industrial applications [27]. Keeping this in mind, we set up lab enrichments on different carbon sources: commercial medium YTS and natural media from energy crops such as pretreated

and woodchips of *A. donax*, woodchips of *C. cardunculus*. Our strategy was based on a metagenomic analysis of the enriched growths to avoid possible limitations due to the difficult isolation of archaeal strains.

It was expected that enrichments on recalcitrant biomasses could lead to the loss of major portions of the microbial population present in environmental samples [43]. Thus, to maintain the highest microbial diversity in the samples during the enrichment, a mild selective pressure for a short period was applied. More specifically, we used the growth media on which the genera, identified in the previous study [27] demonstrated the ability to grow [44–47]. Next, during the cultivation, the carbon source was slowly replaced by adding natural biomasses.

After lab enrichments, a remarkable reduction of the microbial diversity and a switch in the composition of the microbial population were observed. A dramatic decrease in the number of identified ORFs, passing from 14,934 in Pool1 [27] to a range between 9947–4300 in the enriched samples (Table S1), confirmed the strong selection of the microorganisms present in Pool1 after the enrichments. Moreover, Pool1 was dominated by *Acidianus hospitalis*, representing more than 50% of the entire microbiome [27] (Figure 1) while the enriched population was dominated by the genus *Saccharolobus* (mainly *S. solfataricus*), followed by *Sulfolobus* (which was only ~1.5% in Pool1) and by not assigned microorganisms (Figures 1 and S2). This might reflect the ability of *S. solfataricus* to grow aerobically that might have outcompeted *A. hospitalis* [48]. Another strong selection might have occurred because of the carbon source used for the enrichment and the growing medium. *S. solfataricus* can grow on a variety of sugars, such as polysaccharides (cellulose, starch, dextrin), disaccharides (maltose, sucrose, lactose), hexoses (e.g., D-galactose, D-glucose, D-mannose and L-fucose), and pentoses (e.g., D-xylose and L-arabinose) [49], reflecting the abundance of CAZymes in *S. solfataricus* P2 (28 GHs belonging to 16 different families). In contrast, *A. hospitalis* shows only 19 GHs belonging to 10 different families [28] and, although being able to grow on yeast extract [50], the use of Brock medium and the aerobic conditions might have limited its growth on YTS medium.

The functional annotation of the ORFs identified in the analyzed samples revealed 267 CAZymes in the YTS medium sample, 208 in Arundo woodchip, 194 in Arundo pretreated, and 135 in Thistle woodchip representing ~3% of the total ORFs. Among the CAZymes, 312 sequences have been annotated as putative GHs belonging to 23 different families (Figure 2), 83 from the Arundo woodchip sample, 78 from the Arundo pretreated and 52 from the Thistle woodchip.

It is worth noting that the growth on media containing YTS and Arundo and Thistle biomasses resulted in a stronger selection toward CAZymes active on substrates containing axial rather than equatorial C1-OR bonds. In fact, the number of sequences for CAZymes active on the latter substrates in microbiomes obtained from the enrichments was similar and in one case much reduced if compared to the CAZymes' sequences composition of Pool1 (Figure 2). Remarkably, the number of GH1 sequences is halved after the enrichments. This can be explained from the observation that Pool1 in solfatara Pisciarelli is surrounded by a rich and diverse vegetation that, dropping in the pool, might be a source of different (hemi)celluloses. CAZymes belonging to GH1 show several substrate specificities, including glycosides of glucose, galactose, xylose, mannose, glucuronic acid, fucose, phospho-glucose, and phospho-galactose all involved in  $\beta$ -O-glycosidic bonds frequently found in the plant (hemi)celluloses [42]. Enrichments on Arundo and Thistle biomasses, might have selected for GH1 CAZymes specific for the lignocellulose present in these sources.

A different trend was observed for families grouping enzymes active on axial C1-OR. A substantial increase in number was observed among sequences belonging to GH15, GH31, GH122, and GH133, and, remarkably, we could identify several sequences from GH109, a family not found in Pool1. Possibly, starch is more abundant in Arundo [23] than in the plant biomass present in Pisciarelli solfatara [27]. Alternatively, starch is more persistent at the conditions used for the enrichment if compared to those observed in vivo at very low



pH and high temperatures. Therefore, specific CAZymes might have been selected for its hydrolysis and acted more efficiently than (hemi)cellulases on the lignocellulose component of Arundo and Thistle. However, it is worth noting that starch is a relevant component of these energy crops, which are often considered mainly as sources of lignocellulose.

The sequences annotated as putative GHs showed a wide range of identities towards the characterized GHs present in the CAZy database ([www.cazy.org](http://www.cazy.org)). In particular, although 19% have 100% identity with already characterized sequences, more than 50% of the sequences, identified in the enriched samples, represent completely new GHs showing identity  $\leq 60\%$  towards those characterized (Figure S10). Moreover, about 10% of the hypothetical GHs identified show an identity of 80% and 90%, respectively, compared to those characterized. This is still interesting, as is well-known that an amino acid sequence identity difference of even just 10% could entail substantial differences in terms of substrate specificity, stability, pH dependence, etc. [51–54]. Among this group of putative GHs, a sequence encoding a new uncharacterized archaeal enzyme, homolog to  $\alpha$ -glucosidases belonging to GH31, has been identified. Although a high number of  $\alpha$ -glucosidases belonging to the GH31 family have been isolated from Bacteria and Eukaryota, relatively few are known from Archaea. To date, only six  $\alpha$ -glucosidases (EC 3.2.1.20) from thermophilic archaea have been characterized: AglA (PTO0092) from *Picrophilus torridus* DSM 9790 [37], MaLA (Sso1\_0793) from *Saccharolobus solfataricus* 98/2 [33], MaLA (Saci\_1160) from *Sulfolobus acidocaldarius* DSM639 [36], MaLA (SSO3051) from *S. solfataricus* P2 [39], ST2525 from *S. tokodaii* str.7 [35], and AglA (Ta0298) from *Thermoplasma acidophilum* DSM 1728 [38]. The predicted product of the gene *aglu1* showed a sequence identity of 88% with a putative  $\alpha$ -glucosidase of the hyperthermophilic crenarchaeon *Saccharolobus shibatae* (WP\_240781539.1) and 82% with the characterized  $\alpha$ -glucosidase from the hyperthermophilic crenarchaeon *S. solfataricus* MaLA (SSO3051). It is worth noting that the analysis of the metagenomic contig containing *aglu1* gene, as well as four ORFs related to the maltose transport system MalK/MalG/MalF (Figure S4b), demonstrated its phylogenetic distance from the genera *Saccharolobus* and *Sulfolobus*. However, the relationship to the family of *Sulfolobaceae*, with a genomic environment similar to *S. solfataricus* P2 and *S. shibatae*, suggests that it belongs to a chromosome of a new, unclassified genus (Figures 6 and S4).

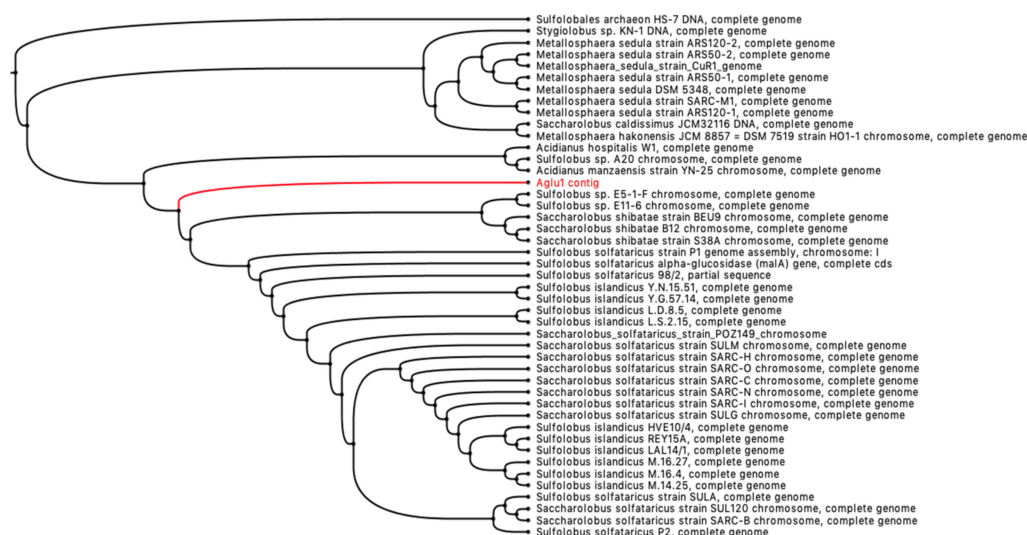


Figure 6. Phylogenetic blastx tree of the metagenomic contig containing the gene *aglu1* (branch in red).

The deduced protein sequence of Aglu1 was multi-aligned toward the archaeal characterized enzymes belonging to GH31, showing the presence of the conserved catalytic residues of this family, and suggesting that the gene encodes for a functional enzyme. Indeed, the recombinant pure Aglu1 was optimally active on 4Np- $\alpha$ -Glc in a sodium ac-

etate buffer in the range of pH 3.5–5.0 (Figure 4a) at 95 °C (Figure 4c) showing remarkable stability at 75 °C, 85 °C and 95 °C (Figure 4d).

Aglu1 showed high activity on both  $\alpha$  1,4- and  $\alpha$  1,2-glycosidic bonds with a similar catalytic efficiency on both maltose and kojibiose (Table 1). It is worth noting that Aglu1, though being classified as a Type II  $\alpha$ -glucosidase, since it hydrolyses preferentially maltose and small maltooligosaccharides over aryl- $\alpha$ -glycosides (Table 1) [55], is the first archaeal GH31 that is reported to be active on kojibiose, showing a catalytic efficiency on this substrate similar to that on maltose.

Compared to other characterized archaeal GH31, Aglu1 was the most active on 4Np- $\alpha$ -Glc (see the  $k_{cat}$  in row 1 of Table 2) and preferentially hydrolyzed the  $\alpha$ -1,4-glycosidic linkages within short chain length. It worth mentioning that  $k_{cat}$  of Aglu1 was 327 fold higher than that of ST2525 enzyme from *S. tokodaii*, although kinetic parameters of ST2525 were measured only at a temperature that was 5 °C lower than Aglu1 (80 °C vs. 85 °C) [35] (Table 2). Moreover, Aglu1 showed a  $k_{cat}$  ~20- fold higher than MalA from *S. acidocaldarius* on 4Np- $\alpha$ -Glc, although kinetic parameters of the latter enzyme were measured at 95 °C, only 5 °C below its optimal temperature [33]. Aglu1 was also the second most active enzyme on maltose and maltotriose after MalA from *S. acidocaldarius* (see the  $k_{cat}$  in row 2 and 3 of Table 2) while it was completely inactive on 4NP- $\alpha$ -Xyl, differently from the GH31 XylS from *S. solfataricus* [34]. For its ability to catalyze the hydrolysis of maltodextrines at a high temperature, Aglu1 could be an interesting candidate to be included in the enzymatic cocktails used for the combined liquefaction/saccharification of starch, a well know application of  $\alpha$ -glucosidases. In addition, its activity on kojibiose, a low caloric disaccharide, may offer the possibility of using Aglu1 synthesizing this sugar by transglycosylation or as a modified *glycosynthase* through an approach combining enzyme and reaction engineering [56–58].

**Table 2.** Comparison of the steady-state kinetic constants of Aglu1 with all archaeal characterized GH31. <sup>a</sup> (This study), <sup>b</sup> [56], <sup>c</sup> [37], <sup>d</sup> [36], <sup>e</sup> [35], <sup>f</sup> [38]. Kinetic Parameters of Aglu1, MalA from *S. solfataricus* and AglA from *P. torridus* were measured at 85 °C, while MalA from *S. acidocaldarius*, ST2525 from *S. tokodaii* and AglA from *T. acidophilum* at 95 °C, 80 °C and 80 °C, respectively. ND: not detected.

Substrate	Kinetic Parameter	Aglu1 <sup>a</sup>	MalA <sup>b</sup> <i>S. solfataricus</i>	AglA <sup>c</sup> <i>P. torridus</i>	MalA <sup>d</sup> <i>S. acidocaldarius</i>	ST2525 <sup>e</sup> <i>S. tokodaii</i>	AglA <sup>f</sup> <i>T. acidophilum</i>
1 4Np- $\alpha$ -Glc	$K_M$ (mM)	3.59 $\pm$ 0.33	1.7 $\pm$ 0.2	0.94 $\pm$ 0.14	0.87 $\pm$ 0.18	3.18	1.77
	$k_{cat}$ (s <sup>-1</sup> )	327.4 $\pm$ 8.3	38.5 $\pm$ 1.6	65.1	16.7	1.05	59.7
	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )	91.14	22.1	69.25	19.2	0.33	33.7
2 Maltose	$K_M$ (mM)	3.16 $\pm$ 0.65	2.00 $\pm$ 0.12 <sup>a</sup>	3.70 $\pm$ 0.32	0.80 $\pm$ 0.32	10.9	0.65
	$k_{cat}$ (s <sup>-1</sup> )	901.6 $\pm$ 53.5	481.1 $\pm$ 6 $10^{-5}$ <sup>a</sup>	188.1	1277.8	6.5	143
	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )	285.0	239.6 <sup>a</sup>	50.83	1597.2	0.598	220
3 Maltotriose	$K_M$ (mM)	1.03 $\pm$ 0.20			0.30 $\pm$ 0.08	3.11	0.47
	$k_{cat}$ (s <sup>-1</sup> )	690.7 $\pm$ 24.4	ND	ND	550	2.53	90.2
	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )	670.64			1833.3	0.814	194

Aglu1 and MalA from *S. solfataricus* shared a high percent of sequence identity (>80%) and the same oligomeric organization into a hexamer. In silico molecular modelling on MalA structure revealed that Aglu1 overall structure is also conserved, except for the domains C and D which were the regions that displayed the lowest identity (50% MalA) (Table S3). As reported in [39], these domains were probably gained early in the evolution of GH31 and subsequently, they diverged considerably by conferring specialized properties and being the main driving force for the different enzymatic activities found in the GH31 family [39]. A detailed structural characterization of Aglu1 goes beyond the aims of this work, but, possibly, the differences observed in domains C and D between Aglu1 and MalA may be responsible for the  $k_{cat}$  that is 10- and 2-fold higher in Aglu1 vs. MalA, on 4Np- $\alpha$ -Glc and maltose, respectively.

In this study, we demonstrated that samples taken from solfataric environments can be used to select a microbial consortium able to grow efficiently in a lab on biomasses from energy crops. Instead of a classical selection of new strains, our metagenomic approach was innovative and efficient. The analysis of the metagenomic data on three enrichments allowed for a global view of the microbial composition of the selected consortia and to identify many novel genes encoding for potential GHs. Surprisingly, the selection produced a number of genes encoding for CAZymes potentially active on starch that was higher than those potentially active on (hemi)cellulose. In addition, the expression and characterization of one of these genes allowed to identify a novel  $\alpha$ -glycosidase, from a novel unknown archaeon, that was the first GH31 able to convert efficiently kojibiose into glucose.

## 4. Materials and Methods

### 4.1. Materials

All commercially available substrates were purchased from Merck, Carbosynth and Megazyme. The synthetic oligonucleotides were from Eurofins (Italy). The pretreated biomass of *A. donax* used in this study was derived from a pretreatment step by PROESA® technology of the Chemtex Group. *C. cardunculus* used in this study was provided by Novamont. Both biomasses were provided in the frame of the project “PON 01\_0966 ENERBIOCHEM”.

### 4.2. Environmental Sampling

The sample from the hydrothermal mud/water Pool1 in Pisciarelli solfatara was collected into sterile bottles and immediately transferred to the laboratory for in vitro cultivation. In situ measurements of temperature and pH were performed by using a HI-93510 thermometer (HANNA instruments, Padova, Italy) equipped with a Pt100 probe and litmus tests. Next, pH was accurately measured again with a pH meter (Crison Instruments, Inc., Barcelona, Spain) in the laboratory.

### 4.3. Microbial Enrichment and Isolation of Metagenomic DNA

For the enrichment setup, three aliquots of the environmental sample were grown in three different basal nutrient media, selected based on microbial consortium populating Pisciarelli solfatara: Basal Salt Medium (BSM) at pH 1.8 [29], Brock salt medium pH 3.5 [30], and Pyrobaculum (PYR) salt medium pH 7.0 [31], supplemented by tryptone, sucrose and yeast extract 0.1% as initial carbon source. Every three days, serial dilution (1:5 *v/v*) of the cultures in fresh salt medium with pretreated *Arundo donax* (0.15% *w/v*) was performed. Contextually, a control culture was performed by using trypton, sucrose and yeast extract 0.1% as carbon source.

For the enrichment experiment, the environmental sample was grown on Brock salt medium pH 3.5 [30] supplemented by tryptone, sucrose and yeast extract 0.1% as initial carbon source. After three days, the culture was split into three aliquots and diluted 1:5 *v/v* in a fresh salt medium with three different biomasses each one: pretreated *A. donax*, woodchips *A. donax*, woodchips *Cynara cardunculus*. Every three days, serial dilution 1:5 *v/v* of the cultures in fresh salt medium with biomasses was performed. Contextually, a control culture was set up by using tryptone, sucrose and yeast extract 0.1% as carbon source.

The enriched microbial populations were recovered by centrifugation at 5000 g for 20 min at RT. The DNA from the enriched microbial population was extracted with Power Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) by following the manufacturer’s protocol.

The extracted and purified DNA from enriched samples was used for shotgun sequencing with HiSeq2000 (Illumina) performed at Beijing Genomics Institute (BGI-Shenzhen) Shenzhen, China.

The sequencing reads are available in the NCBI Sequence Read Archive (SRA) database in the Bioproject PRJNA861126 under the accession numbers SAMN29886219

(YTS), SAMN29933369 (Arundo woodchip), SAMN29933377 (Arundo pretreated), and SAMN29933397 (Thistle woodchips).

#### 4.4. Taxonomic Analysis and Functional Annotation

For microbial diversity analysis, short paired-end Illumina reads (90 bp) of each sample were analyzed by using Kaiju against the NCBI RefSeq protein database [59].

Clean reads were assembled using MEGAHIT [60] in “meta-sensitive” mode by using min-count = 2 and k-mers 21, 31, 41, 51, 61, 71, 81, 91, and 99.

All the contigs obtained by the assembly were analyzed by using Prodigal [61] to identify the open reading frames and analyzed by dbCAN2 pipeline for the carbohydrate-active enzymes annotation [32].

The ORFs assigned as glycosidases were analyzed by using Diamond in *bastp* mode [62], against a custom protein database with the 7190 sequences of the characterized GHs from the CAZy database [28].

The contig encoding Aglu1 and the Aglu1 sequence are available in the GenBank database under the accession numbers OP149530 and OP149529, respectively.

#### 4.5. Cloning, Expression, and Purification of Recombinant Aglu1

The *aglu1* gene was amplified by PCR from the metagenome DNA extracted from enrichment on pretreated *A. donax* using the synthetic oligonucleotides GH31\_forward (5'-GGAATTCATATGCAAGTAACAAAGATATACGAGAG-3') and GH31\_reverse (5'-CCGCTCGAGCCTTCTTAAGTTAATATTTTTATCC-3').

The amplification reaction was performed with the PfuUltra Ultra HF DNA Polymerase (Stratagene) by using the following program: hot start 5 min at 95 °C; 5 cycles 1 min at 95 °C, 1 min at 50 °C and 1.5 min at 72 °C; 30 cycles 1 min at 95 °C, 1 min at 60 °C, and 1.5 min at 72 °C; final extension 10 min at 72 °C. The DNA fragment obtained was cloned in the expression vector pET20b (Invitrogen, Waltham, MA, USA), obtaining the recombinant plasmid pET20b-Aglu1. The PCR-generated construct was verified by sequencing and the ORF was expressed in *Escherichia coli* cells, strain BL21 (DE3) RIL (Invitrogen), according to the manufacturer. The cells transformed with pET20b-Aglu1 were grown at 37 °C in 2 L of Super Broth at 37 °C supplemented with ampicillin (50 µg mL<sup>-1</sup>) and chloramphenicol (30 µg mL<sup>-1</sup>). Gene expression was induced by the addition of 0.5 mM IPTG when the culture reached an A600 of 0.6 OD. Growth was allowed to proceed for 16 h, and cells were harvested by centrifugation at 5000 × *g*. The resulting cell pellet was resuspended in 50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl and 1% TRITON-X100 with a ratio of 3:1 (*v/w*) and then incubated at 37 °C for 1 h with 20 mg of lysozyme (Fluka) and 25 U g<sup>-1</sup> cell of Benzonase (Novagen, Madison, WI, USA). Cells were lysed by French cell pressure treatment and cell debris was removed by centrifugation at 12,000 × *g* for 30 min. The free cellular extract (FCE) was heat-fractionated three times (i) for 30 min at 50 °C, (ii) 30 min at 60 °C, and (iii) for 30 min at 70 °C. The resulting supernatant was loaded on Q Sepharose High Performance 16/10 (GE-Healthcare, Chicago, IL, USA) equilibrated with Tris HCl 20 mM pH 8.0 (Buffer A). After an initial wash-step (2-column volumes, CV) with buffer A, the protein was eluted with a 3-CV gradient up to 100% of Tris HCl 20 mM pH 8.0 + 1 M NaCl (Buffer B), followed by 2-CV of Buffer B at a flow rate of 2 mL min<sup>-1</sup>. The active fractions were pooled and dialyzed against Tris HCl 20 mM pH 8.0. (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 1 M was added to dialysed pool successively loaded on Phenyl Sepharose High Performance 26/10 (GE-Healthcare) equilibrated with Tris HCl 20 mM pH 8.0 + (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 1 M (Buffer A). After an initial wash-step (1.5 CV) with buffer A, the protein was eluted with a 5 CV gradient up to 100% of Tris HCl 20 mM pH 8.0 (Buffer B), followed by 2 CV of Buffer B at a flow rate of 3 mL min<sup>-1</sup>. The active fractions were pooled and dialyzed against Tris-HCl 20 mM pH 8.0 and concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da).

#### 4.6. Molecular Mass Determination

The molecular mass of Aglu1 was determined by gel filtration on a Superose 6 10/300 FPLC column (GE-Healthcare) with an isocratic solution of 20 mM Tris HCl (pH 8.0). Molecular weight markers were Thyroglobulin (669 kDa), apoferritin (443 kDa), Bovin Albumin (BSA) (66 kDa) and Cytochrome C (12.7 kDa).

#### 4.7. Standard Assay

The standard assay for Aglu1 was performed on 5 mM 4Np- $\alpha$ -Glc in 50 mM sodium acetate buffer at pH 4.5 at 65 °C by using 2  $\mu$ g of enzyme in the final volume of 0.2 mL. After 1 min of incubation at 65 °C, the reaction was blocked in ice and by adding 0.8 mL of 1 M sodium carbonate pH 10.2. The absorbance was measured at 420 nm at room temperature; the mM extinction coefficient of 4-nitrophenol in this condition is 17.2 mM<sup>-1</sup> cm<sup>-1</sup>. In all the assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme. One unit of activity was defined as the amount of enzyme catalyzing the conversion of 1  $\mu$ mole of substrate into the product in 1 min, at the conditions described.

#### 4.8. Temperature and pH Influence

The temperature and pH optima were determined by assaying Aglu1 in 50 mM of the indicated buffers at different pHs in the range of 40–100 °C in the standard assay conditions. Thermal stability was evaluated by incubating the enzymes in Tris HCl 20 mM pH 8.0, at the indicated temperatures (75 °C, 85 °C, 95 °C). At intervals, aliquots (2  $\mu$ g of enzyme) were withdrawn, transferred in ice, centrifuged for 1 min at 16,000 $\times$  *g* and assayed at standard conditions. The residual activities were expressed as a percentage of the maximal enzymatic activity measured before the incubation at indicated temperatures.

All the experiments were performed in duplicate.

#### 4.9. Substrate Specificity and Steady-State Kinetic Constants

The activity of Aglu1 was tested on different aryl-glycosides, oligosaccharides (maltose, maltotriose, maltotetraose, maltopentaose, trehalose, and nigerose), and polysaccharides (starch, and glycogen) in 50 mM sodium acetate buffer at pH 4.5 and 65 °C for 1 h. D-Glucose released by the hydrolysis of oligo- and polysaccharides can be measured by using Megazyme D-Glucose GOPOD kit according to manufacturing protocol.

To define the mode of action (*endo*- versus *exo*-acting), Aglu1 was tested on maltodextrins ranging from two to five glucose residues. The products were analyzed on silica gel 60 F254 TLC by using acetone/isopropanol/water (60:30:15 v:v) as eluent and 5% sulfuric acid in methanol for detection.

Kinetic constants of Aglu1 on 4Np- $\alpha$ -Glc and maltose were measured in 50 mM sodium acetate buffer at pH 4.5 at 85 °C by using substrates ranging from 0.5 to 35 mM and from 0.2 to 30 mM, respectively. Spontaneous hydrolysis of the substrates was subtracted by using appropriate blank mixtures without enzyme.

Kinetic constants of MalA on maltose were measured in 100 mM sodium citrate buffer at pH 5.0 at 85 °C. Spontaneous hydrolysis of the substrates was subtracted by using appropriate blank mixtures without enzyme. All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program Prism 5.0 (GraphPad Software, San Diego, CA, USA).

## 5. Conclusions

The metagenomic analysis of cultures of hyperthermophilic microorganisms enriched on plant biomasses allowed the identification of several genes encoding for GHs still unclassified and uncharacterized, and potentially useful for biotechnological application. The use of vegetable wastes as culture media for consortia of extremophilic microbes can represent a cheap carbon source to achieve two promising targets: a method of waste

handling at lower environmental impact and the possibility to select extremozymes of biotechnological interest using cheap feedstocks not competing with food supplies.

Indeed, enrichments on two specific energy crops showed that a remarkable selection toward a high number of diverse carbohydrate active enzymes. Surprisingly, the selection was more efficient on starch-degrading enzymes rather than on biocatalysts specific on lignocellulose. The enrichments allowed us to identify and characterize a novel, highly thermophilic and thermostable archaeal  $\alpha$ -glucosidase, whose genetic organization suggests that it belongs to a new member of the *Sulfolobaceae*. Aglu1 hydrolyses maltose and small maltooligosaccharides with specificity constants higher than those observed on aryl-glucosides, and an unexpected specificity on kojibiose. The enzymatic characterization demonstrated that Aglu1 is the archaeal GH31 most active on 4Np- $\alpha$ -Glc and one of the most active enzymes from this source on maltose.

This study demonstrated that in-lab enrichments of microbial communities from extreme environments represent an efficient strategy for the identification of novel thermophilic and thermostable CAZymes and that energy crops work as rich media allowing the identification of enzymes able to act cooperatively to produce fermentable sugars not only from lignocellulose, but also from starch.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231810505/s1>.

**Author Contributions:** R.I., A.S. and M.M., conceptualization of the study; R.I., A.S. and B.C.-P., performed the data curation; R.I., A.S., B.C.-P. and M.M., performed formal analysis of the data; R.I., R.G. and F.B., performed investigations; M.M., collected funding resources; B.C.-P. and M.M., performed supervision; R.I., A.S. and M.M., wrote the original draft; R.I.; A.S.; B.C.-P. and M.M., performed review and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grant from the Ministero dell'Università e della Ricerca Scientifica—Industrial Research Project “Integrated agro-industrial chains with high energy efficiency for the development of eco-compatible processes of energy and biochemicals production from renewable sources and for the land valorization (Enerbio-Chem)” PON01 01966, funded in the frame of Operative National Programme Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010. The funding source did not have any role in the study design.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank Francesca Maria Pia Paragliola for performing the kinetic parameters measurements of MalA and Katherine Anne Hominuk for English language editing.

**Conflicts of Interest:** The authors declare they have no conflict of interest.

## References

1. Cobucci-Ponzano, B.; Strazzulli, A.; Iacono, R.; Masturzo, G.; Giglio, R.; Rossi, M.; Moracci, M. Novel thermophilic hemicellulases for the conversion of lignocellulose for second generation biorefineries. *Enzym. Microb. Technol.* **2015**, *78*, 63–73. [CrossRef]
2. Curci, N.; Strazzulli, A.; De Lise, F.; Iacono, R.; Maurelli, L.; Dal Piaz, F.; Cobucci-Ponzano, B.; Moracci, M. Identification of a novel esterase from the thermophilic bacterium *Geobacillus thermodenitrificans* NG80-2. *Extremophiles* **2019**, *23*, 407–419. [CrossRef]
3. Dumorne, K.; Cordova, D.C.; Astorga-Elo, M.; Renganathan, P. Extremozymes: A Potential Source for Industrial Applications. *J. Microbiol. Biotechnol.* **2017**, *27*, 649–659. [CrossRef] [PubMed]
4. Zhu, D.; Adebisi, W.A.; Ahmad, F.; Sethupathy, S.; Danso, B.; Sun, J. Recent Development of Extremophilic Bacteria and Their Application in Biorefinery. *Front. Bioeng. Biotechnol.* **2020**, *8*, 483. [CrossRef] [PubMed]
5. Saini, J.K.; Saini, R.; Tewari, L. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. *3 Biotech* **2015**, *5*, 337–353. [CrossRef] [PubMed]
6. Sadh, P.K.; Duhan, S.; Duhan, J.S. Agro-industrial wastes and their utilization using solid state fermentation: A review. *Bioresour. Bioprocess.* **2018**, *5*, 1–15. [CrossRef]

7. Sarsaiya, S.; Jain, A.; Kumar Awasthi, S.; Duan, Y.; Kumar Awasthi, M.; Shi, J. Microbial dynamics for lignocellulosic waste bioconversion and its importance with modern circular economy, challenges and future perspectives. *Bioresour. Technol.* **2019**, *291*, 121905. [[CrossRef](#)]
8. Di Donato, P.; Fiorentino, G.; Anzelmo, G.; Tommonaro, G.; Nicolaus, B.; Poli, A. Re-Use of Vegetable Wastes as Cheap Substrates for Extremophile Biomass Production. *Waste Biomass Valorization* **2011**, *2*, 103–111. [[CrossRef](#)]
9. Robak, K.; Balcerek, M. Current state-of-the-art in ethanol production from lignocellulosic feedstocks. *Microbiol. Res.* **2020**, *240*, 126534. [[CrossRef](#)]
10. De Medeiros, T.D.M.; Dufossé, L.; Bicas, J.L. Lignocellulosic substrates as starting materials for the production of bioactive biopigments. *Food Chem.* **2022**, *13*, 100223. [[CrossRef](#)]
11. Grewal, J.; Khare, S.K. One-pot bioprocess for lactic acid production from lignocellulosic agro-wastes by using ionic liquid stable *Lactobacillus brevis*. *Bioresour. Technol.* **2018**, *251*, 268–273. [[CrossRef](#)] [[PubMed](#)]
12. Jiménez-Quero, A.; Pollet, E.; Avérous, L.; Phalip, V. Optimized Bioproduction of Itaconic and Fumaric Acids Based on Solid-State Fermentation of Lignocellulosic Biomass. *Molecules* **2020**, *25*, 1070. [[CrossRef](#)] [[PubMed](#)]
13. Namnuch, N.T.A.; Thammasittirong, S.N.R. Lignocellulose hydrolytic enzymes production by *Aspergillus flavus* KUB2 using submerged fermentation of sugarcane bagasse waste. *Mycology* **2021**, *12*, 119–127. [[CrossRef](#)] [[PubMed](#)]
14. De Oliveira Rodrigues, P.G.L.; Pasquini, D.; Badotti, F.; Góes-Neto, A.; Baffi, M.A. Lignocellulose-degrading enzymes production by solid-state fermentation through fungal consortium among Ascomycetes and Basidiomycetes. *Renew. Energy* **2020**, *145*, 2683–2693. [[CrossRef](#)]
15. Kawaguchi, H.; Ogino, C.; Kondo, A. Microbial conversion of biomass into bio-based polymers. *Bioresour. Technol.* **2017**, *245*, 1664–1673. [[CrossRef](#)]
16. Patel, S.K.S.; Gupta, R.K.; Das, D.; Lee, J.-K.; Kalia, V.C. Continuous biohydrogen production from poplar biomass hydrolysate by a defined bacterial mixture immobilized on lignocellulosic materials under non-sterile conditions. *J. Clean. Prod.* **2021**, *287*, 125037. [[CrossRef](#)]
17. Fernando Sandun, A.S. Chandrapal Chauda, and Murali Naveen, Biorefineries: Current Status, Challenges, and Future Direction. *Energy Fuels* **2006**, *20*, 1727–1737. [[CrossRef](#)]
18. Kamm, B.; Kamm, M. Principles of biorefineries. *Appl. Microbiol. Biotechnol.* **2004**, *64*, 137–145. [[CrossRef](#)]
19. Ilari, A.; Duca, D.; Boakye-Yiadom, K.A.; Gasperini, T.; Toscano, G. Carbon Footprint and Feedstock Quality of a Real Biomass Power Plant Fed with Forestry and Agricultural Residues. *Resources* **2022**, *11*, 7. [[CrossRef](#)]
20. Shatalov, A.A.; Pereira, H. Dissolving grade eco-clean cellulose pulps by integrated fractionation of cardoon (*Cynara cardunculus* L.) stalk biomass. *Chem. Eng. Res. Des.* **2014**, *92*, 2640–2648. [[CrossRef](#)]
21. Branca, F.; Argento, S.; Paoletti, A.M.; Melilli, M.G. The Physiological Role of Inulin in Wild Cardoon (*Cynara cardunculus* L. var. *sylvestris* Lam.). *Agronomy* **2022**, *12*, 290. [[CrossRef](#)]
22. Lemons e Silva, C.F.; Schirmer, M.A.; Maeda, R.N.; Barcelos, C.A.; Pereira, N. Potential of giant reed (*Arundo donax* L.) for second generation ethanol production. *Electron. J. Biotechnol.* **2015**, *18*, 10–15. [[CrossRef](#)]
23. Proietti, S.; Moscatello, S.; Fagnano, M.; Fiorentino, N.; Impagliazzo, A.; Battistelli, A. Chemical composition and yield of rhizome biomass of *Arundo donax* L. grown for biorefinery in the Mediterranean environment. *Biomass Bioenergy* **2017**, *107*, 191–197. [[CrossRef](#)]
24. Juerges, N.; Hansjürgens, B. Soil governance in the transition towards a sustainable bioeconomy—A review. *J. Clean. Prod.* **2018**, *170*, 1628–1639. [[CrossRef](#)]
25. Strazzulli, A.; Iacono, R.; Giglio, R.; Moracci, M.; Cobucci-Ponzano, B. Hyperthermophilic Environments: Biodiversity and Biotechnology. In *Microbial Ecology of Extreme Environments*; Chénard, C.L.F., Lauro, F.M., Eds.; Springer: Cham, Switzerland, 2017.
26. Iacono, R.; Cobucci-Ponzano, B.; De Lise, F.; Curci, N.; Aurelli, L.; Moracci, M.; Strazzulli, A. Spatial Metagenomics of Three Geothermal Sites in Pisciarelli Hot Spring Focusing on the Biochemical Resources of the Microbial Consortia. *Molecules* **2020**, *25*, 4023. [[CrossRef](#)]
27. Strazzulli, A.; Cobucci-Ponzano, B.; Iacono, R.; Giglio, R.; Aurelli, L.; Curci, N.; Schiano-di-Cola, C.; Santangelo, A.; Contursi, P.; Lombard, V.; et al. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* **2020**, *287*, 1116–1137. [[CrossRef](#)]
28. Drula, E.; Garron, M.L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Res.* **2022**, *50*, D571–D577. [[CrossRef](#)]
29. Shiers, D.W.; Blight, K.R.; Ralph, D.E. Sodium sulphate and sodium chloride effects on batch culture of iron oxidising bacteria. *Hydrometallurgy* **2005**, *80*, 75–82. [[CrossRef](#)]
30. Brock, T.D.; Brock, K.M.; Belly, R.T.; Weiss, R.L. *Sulfobolus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **1972**, *84*, 54–68. [[CrossRef](#)]
31. Amo, T.; Paje, M.L.; Inagaki, A.; Ezaki, S.; Atomi, H.; Imanaka, T. *Pyrobaculum calidifontis* sp. nov., a novel hyperthermophilic archaeon that grows in atmospheric air. *Archaea* **2002**, *1*, 113–121. [[CrossRef](#)]
32. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [[CrossRef](#)] [[PubMed](#)]

33. Rolfsmeier, M.; Blum, P. Purification and characterization of a maltase from the extremely thermophilic crenarchaeote *Sulfolobus solfataricus*. *J. Bacteriol.* **1995**, *177*, 482–485. [[CrossRef](#)] [[PubMed](#)]
34. Moracci, M.; Cobucci Ponzano, B.; Trincone, A.; Fusco, S.; De Rosa, M.; van Der Oost, J.; Sensen, C.W.; Charlebois, R.L.; Rossi, M. Identification and molecular characterization of the first alpha-xylosidase from an archaeon. *J. Biol. Chem.* **2000**, *275*, 22082–22089. [[CrossRef](#)] [[PubMed](#)]
35. Park, J.E.; Park, S.H.; Woo, J.Y.; Hwang, H.S.; Cha, J.; Lee, H. Enzymatic properties of a thermostable alpha-glucosidase from acidothermophilic crenarchaeon *Sulfolobus tokodaii* strain 7. *J. Microbiol. Biotechnol.* **2013**, *23*, 56–63. [[CrossRef](#)]
36. Choi, K.H.; Hwang, S.; Cha, J. Identification and characterization of MalA in the maltose/maltodextrin operon of *Sulfolobus acidocaldarius* DSM639. *J. Bacteriol.* **2013**, *195*, 1789–1799. [[CrossRef](#)]
37. Angelov, A.; Putyrski, M.; Liebl, W. Molecular and biochemical characterization of alpha-glucosidase and alpha-mannosidase and their clustered genes from the thermoacidophilic archaeon *Picrophilus torridus*. *J. Bacteriol.* **2006**, *188*, 7123–7131. [[CrossRef](#)]
38. Seo, S.-H.; Choi, K.-H.; Hwang, S.; Kim, J.; Park, C.-S.; Rho, J.-R.; Cha, J. Characterization of the catalytic and kinetic properties of a thermostable *Thermoplasma acidophilum*  $\alpha$ -glucosidase and its transglucosylation reaction with arbutin. *J. Mol. Catal. B Enzym.* **2011**, *72*, 305–312. [[CrossRef](#)]
39. Ernst, H.A.; Lo Leggio, L.; Willemoes, M.; Leonard, G.; Blum, P.; Larsen, S. Structure of the *Sulfolobus solfataricus* alpha-glucosidase: Implications for domain conservation and substrate recognition in GH31. *J. Mol. Biol.* **2006**, *358*, 1106–1124. [[CrossRef](#)]
40. Mirdita, M.; Schutze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. [[CrossRef](#)]
41. Nakai Hiroyuki, I.T.; Shigeki, T.; Kazuki, M.; Takeshi, Y.; Masayuki, O.; Haruhide, M.; Seiya, C.; Yoshio, S.; Atsuo, K. Plant  $\alpha$ -Glucosidase: Molecular Analysis of Rice  $\alpha$ -Glucosidase and Degradation Mechanism of Starch Granules in Germination Stage. *J. Appl. Glycosci.* **2006**, *53*, 137–142. [[CrossRef](#)]
42. Cobucci-Ponzano, B.; Aurilia, V.; Riccio, G.; Henrissat, B.; Coutinho, P.M.; Strazzulli, A.; Padula, A.; Corsaro, M.M.; Pieretti, G.; Pocsfalvi, G.; et al. A new archaeal beta-glycosidase from *Sulfolobus solfataricus*: Seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J. Biol. Chem.* **2010**, *285*, 20691–20703. [[CrossRef](#)]
43. Entcheva, P.; Liebl, W.; Johann, A.; Hartsch, T.; Streit, W.R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* **2001**, *67*, 89–99. [[CrossRef](#)]
44. Giaveno, M.A.; Urbieta, M.S.; Ulloa, J.R.; Toril, E.G.; Donati, E.R. Physiologic versatility and growth flexibility as the main characteristics of a novel thermoacidophilic *Acidianus* strain isolated from Copahue geothermal area in Argentina. *Microb. Ecol.* **2013**, *65*, 336–346. [[CrossRef](#)]
45. Huber, G.; Spinnler, C.; Gambacorta, A.; Stetter, K.O. *Metallosphaera sedula* gen. and sp. nov. Represents a New Genus of Aerobic, Metal-Mobilizing, Thermoacidophilic Archaeobacteria. *Syst. Appl. Microbiol.* **1989**, *12*, 38–47. [[CrossRef](#)]
46. Volkl, P.; Huber, R.; Drobner, E.; Rachel, R.; Burggraf, S.; Trincone, A.; Stetter, K.O. *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl. Environ. Microbiol.* **1993**, *59*, 2918–2926. [[CrossRef](#)]
47. Zillig, W.; Stetter, K.O.; Wunderl, S.; Schulz, W.; Priess, H.; Scholz, I. The *Sulfolobus*-“*Caldariella*” group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **1980**, *125*, 259–269. [[CrossRef](#)]
48. You, X.Y.; Liu, C.; Wang, S.Y.; Jiang, C.Y.; Shah, S.A.; Prangishvili, D.; She, Q.; Liu, S.J.; Garrett, R.A. Genomic analysis of *Acidianus hospitalis* W1 a host for studying crenarchaeal virus and plasmid life cycles. *Extremophiles* **2011**, *15*, 487–497. [[CrossRef](#)]
49. Sakai, H.D.; Kurosawa, N. *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. and *Sulfolobus shibatae* as *Saccharolobus shibatae* comb. nov. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1271–1278.
50. Rachel, R.; Bettstetter, M.; Hedlund, B.P.; Haring, M.; Kessler, A.; Stetter, K.O.; Prangishvili, D. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch. Virol.* **2002**, *147*, 2419–2429. [[CrossRef](#)]
51. Addou, S.; Rentsch, R.; Lee, D.; Orengo, C.A. Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *J. Mol. Biol.* **2009**, *387*, 416–430. [[CrossRef](#)]
52. Rost, B.; Liu, J.; Nair, R.; Wrzeszczynski, K.O.; Ofra, Y. Automatic prediction of protein function. *Cell. Mol. Life Sci.* **2003**, *60*, 2637–2650.
53. Tian, W.; Skolnick, J. How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol.* **2003**, *333*, 863–882. [[CrossRef](#)]
54. Todd, A.E.; Orengo, C.A.; Thornton, J.M. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **2001**, *307*, 1113–1143. [[CrossRef](#)]
55. Dusan, V.; Nenad, M.; Dejan, B.; Filip, B.; Segal, A.M.; Dejan, S.; Jovana, T.; Aleksandra, D. The specificity of alpha-glucosidase from *Saccharomyces cerevisiae* differs depending on the type of reaction: Hydrolysis versus transglucosylation. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 6317–6328. [[CrossRef](#)]
56. Yamamoto, K.; Davis, B.G. Creation of an alpha-mannosynthase from a broad glycosidase scaffold. *Angew. Chem. Int. Ed.* **2012**, *51*, 7449–7453. [[CrossRef](#)]
57. Tshililo, N.O.; Strazzulli, A.; Cobucci-Ponzano, B.; Maurelli, L.; Iacono, R.; Bedini, E.; Corsaro, M.M.; Strauss, E.; Moracci, M. The  $\alpha$ -Thioglycoligase Derived from a GH89  $\alpha$ -N-Acetylglucosaminidase Synthesises  $\alpha$ -N-Acetylglucosamine-Based Glycosides of Biomedical Interest. *Adv. Synth. Catal.* **2017**, *359*, 663–676. [[CrossRef](#)]



58. Cobucci-Ponzano, B.; Perugino, G.; Strazzulli, A.; Rossi, M.; Moracci, M. Thermophilic glycosynthases for oligosaccharides synthesis. *Methods Enzymol.* **2012**, *510*, 273–300.
59. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257. [[CrossRef](#)] [[PubMed](#)]
60. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [[CrossRef](#)] [[PubMed](#)]
61. Hyatt, D.; Chen, G.L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)] [[PubMed](#)]
62. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]