



OPEN

# Rapid genome resequencing of an atoxigenic strain of *Aspergillus carbonarius*

SUBJECT AREAS:

GENE REGULATORY  
NETWORKS

FUNGAL GENOMICS

Received

1 August 2014

Accepted

18 February 2015

Published

13 March 2015

Correspondence and  
requests for materials  
should be addressed toF.J.C. (javier.  
cabanes@uab.es)F. Javier Cabañes<sup>1</sup>, Walter Sanseverino<sup>2</sup>, Gemma Castellà<sup>1</sup>, M. Rosa Bragulat<sup>1</sup>, Riccardo Aiese Cigliano<sup>2</sup> & Armand Sánchez<sup>3</sup><sup>1</sup>Veterinary Mycology Group, Department of Animal Health and Anatomy, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain, <sup>2</sup>Sequentia Biotech SL, Barcelona, Catalonia, Spain, <sup>3</sup>Departament de Genètica Animal, Centre de Recerca en AgriGenòmica (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain.

In microorganisms, Ion Torrent sequencing technology has been proved to be useful in whole-genome sequencing of bacterial genomes (5 Mbp). In our study, for the first time we used this technology to perform a resequencing approach in a whole fungal genome (36 Mbp), a non-ochratoxin A producing strain of *Aspergillus carbonarius*. Ochratoxin A (OTA) is a potent nephrotoxin which is found mainly in cereals and their products, but it also occurs in a variety of common foods and beverages. Due to the fact that this strain does not produce OTA, we focused some of the bioinformatics analyses in genes involved in OTA biosynthesis, using a reference genome of an OTA producing strain of the same species. This study revealed that in the atoxigenic strain there is a high accumulation of nonsense and missense mutations in several genes. Importantly, a two fold increase in gene mutation ratio was observed in PKS and NRPS encoding genes which are suggested to be involved in OTA biosynthesis.

Ochratoxin A (OTA) is a potent nephrotoxin which is found mainly in cereals and their products, but it also occurs in a variety of common foods and beverages such chocolate, dried fruits, coffee or wine. This mycotoxin is produced by several species of *Penicillium* and *Aspergillus* among which *Aspergillus carbonarius* is the main responsible source of this mycotoxin in wine, grapes and dried vine fruits from main viticultural regions worldwide<sup>1</sup>.

The biosynthetic pathway of other mycotoxins has been elucidated, as it is the case of the aflatoxin biosynthetic gene cluster of *Aspergillus flavus*, largely responsible for the aflatoxin contamination of agricultural crops<sup>2</sup>. However, little is known about the genes involved in OTA biosynthesis. OTA consists of a polyketide derived chlorinated-dihydromethyl-isocoumarin linked to phenylalanine. Different polyketide synthase (PKS) encoding genes involved in OTA biosynthesis have been identified in various ochratoxin A producing species<sup>3-7</sup>. To date, only a PKS gene involved in the initial steps of the OTA biosynthetic pathway and a nonribosomal peptide synthetase (NRPS) gene have been related to this biosynthetic gene cluster in *A. carbonarius*<sup>8,9</sup>. Recently, the genome sequence of an OTA producer strain of *A. carbonarius* ITEM 5010 (Acv3) has been generated by the United States Department of Energy's Joint Genome Institute using the 454 and the Sanger sequencing technologies (<http://jgi.doe.gov/carbonarius/>)<sup>10</sup>. Similar to many other ascomycetes so far sequenced, the *A. carbonarius* genome contains a large number of PKS and NRPS encoding genes. These genes encode complex and multifunctional proteins involved in the biosynthesis of most of the fungal secondary metabolites. Sequencing technologies may ease the study of the genetic biosynthetic pathways of important mycotoxins, such as OTA. Ion Torrent sequencing technology has been proved to be useful in whole-genome sequencing of small sized genomes, up to 5 Mbp, such as some bacterial genomes<sup>11,12</sup>.

In this study, for the first time, we used the Ion Torrent technology to resequence the genome of a mold, an atoxigenic wild strain of *A. carbonarius* with a genome size of about 36 Mbp, using the reference genome Acv3 of a toxigenic *A. carbonarius* strain. Besides this main objective, and due to the fact that this strain does not produce OTA, we focused some of the bioinformatics analyses in genes involved in OTA biosynthesis.



## Results

**Detection of the OTA production in the *A. carbonarius* strains studied.** *Aspergillus carbonarius* strains grew in CYA medium at 15°C, 25°C and 30°C. Both strains presented good growth with proper sporulation forming typical black colonies (Figure S1). However, the non-OTA producing strain (A-2160) was not able to produce OTA at these temperatures after 7 or 15 days of incubation. On the contrary, *A. carbonarius* ITEM 5010 produced OTA at detectable levels at the three temperatures of incubation tested. This strain produced higher amounts of OTA at 15°C after 15 days of incubation than in the rest of the temperatures and incubation times tested. Figure 1 shows some selected chromatograms of the fungal strains analysed in this study with HPLC. Extracts of *A. carbonarius* ITEM 5010 (Figure 1a) presented a clear peak with the same retention time of OTA (4.7 minutes). The extracts of *A. carbonarius* A-2160 (Figure 1b) showed no signals at the same retention time of OTA.

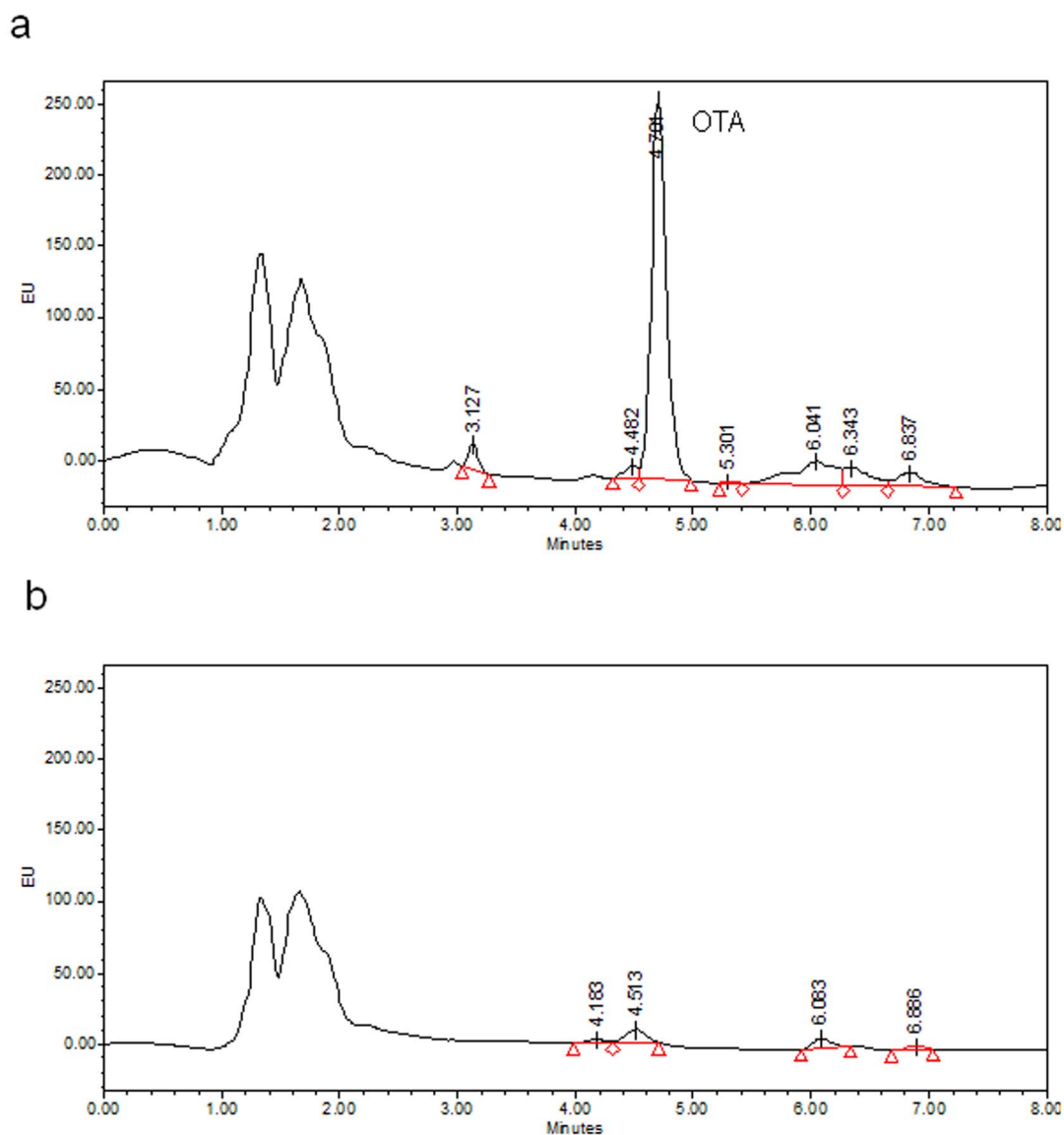
**Resequencing study.** The general resequencing genome data of the atoxigenic strain A-2160 of *A. carbonarius* is summarized in Table 1. A single Ion Torrent run was performed to resequence the genome of

this strain. A total of 3,158,689 reads were sequenced for a total of 547.56 Mbp. The alignment of the reads to the Acv3 showed that more than the 84% of the no-repetitive reference genome was covered with an average depth of  $\sim 12\times$ .

The comparisons between the multiple and unique mapping reads showed that the two strains share about 3 Mb of repetitive regions and that have a similarity higher than 97.2%. The deletions accumulated in the atoxigenic strain of *A. carbonarius* summed to  $\sim 1$  Mbp and the total size of the unmapped reads was of  $\sim 39$  Mb (7.75% of the reads produced by the Ion Torrent run). The alignment between the reads produced from the atoxigenic strain and the Acv3 were used to study the variation between the two strains and to highlight the genomics differences that might explain the phenotype of the atoxigenic strain.

### Single-nucleotide and deletion-insertion polymorphisms analyses.

All uniquely mapped regions were used to perform a specific single-nucleotide polymorphisms (SNPs) and deletion-insertion polymorphisms (DIPs) calling. A total of 52661 high quality SNPs and 7567 high quality DIPs have been identified. Due to the homozygous nature of the re-sequenced genome, all the heterozygous mutations



**Figure 1** | Selected chromatograms of fungal extracts analysed using HPLC coupled to a fluorescence detector of (a) the OTA producing strain of *A. carbonarius* ITEM 5010 (OTA retention time: 4.701 min) and (b) the atoxigenic strain of *A. carbonarius* A-2160, after incubation at 15°C for 7 days on Czapek Yeast extract Agar.


**Table 1 |** Resequencing statistics of the atoxigenic strain of *A. carbonarius*

Total Number of Bases	547.56 Mbp
Number of Q20 Bases	405.83 Mbp
Mean length	173 bp
Longest Read	396 bp
Mean Coverage Depth (Q20)	11.36x
Total Number of reads	5,184,113
Filtered Polyclonal	1,661,202
Filtered Primer Dimer	58
Filtered Low Quality	364,164
Total Number of reads	3,158,689

have been discarded. High quality SNPs and DIPs have been used to perform a mutation mining according with the information of the gene annotation of the reference genome Acv3. Table 2 shows the presence of 7183 missense, 88 nonsense and 528 frameshift mutations. The missense mutations affect 3880 genes. A total of 43 Gene Ontology (GO) families have more than the 5% of mutated genes (Figure 2 and Supplementary Table S1).

**Copy number variation analysis.** Using a read depth approach on the multiple mapping alignment, several copy number variation (CNV) regions were identified. The analysis showed that there were not significant duplicated regions but there were 55 deleted regions (Supplementary Table S2) containing 291 genes that fall in 60 GO classes (Supplementary Table S3). Of that, kinase regulator activity (26.57), serine-type carboxypeptidase activity (10.63) and metal cluster binding (7.97) classes have a high enrichment level. Several small scaffolds in the reference genome Acv3 resulted completely deleted.

**PKS and NRPS encoding gene analyses.** A total of 24 NRPS genes and 25 PKS genes were retrieved with an approach based on Blast, Inteprosca and AntiSmash. For both families a phylogenetic approach was combined with the SNPs, DIPs and CNVs analyses in order to extract supplementary information about these two gene families (Figures 3 and 4). Many NRPS and PKS genes had mutations and one NRPS gene (estExt\_fgenes2\_pg.C\_3\_t20159) and one PKS gene (estExt\_Genemark1.C\_50656) were deleted in the atoxigenic strain.

By using the SNPs information, we calculated the gene mutation ratio (GMR) of the entire set of genes and we compared it with the GMR of the NRPS and PKS gene families. The global GMR was of 1.1 (1 SNP each 909 bp) while the GMR of NRPS and PKS was of 2.2 (1 SNP every 454 bp) and 1.6 (1 SNP every 625 bp), respectively. Moreover the phylogenetic analysis of the two gene families showed that the nearest gene (estExt\_fgenes2\_pg.C\_2\_t10203) to the functional NRPS gene described by Gallo et al.<sup>8</sup> (estExt\_Genemark1.C\_120304: PI 132610) was the most mutated gene in the genome with 201 mutated nucleotides (Figure 3).

In order to confirm some of these mutations by PCR techniques we focused our study in one NRPS gene (PI 132610) and one PKS

gene (PI 173482) that have been reported to be involved in the OTA biosynthetic pathway of *A. carbonarius*<sup>8,9</sup>. In the NRPS gene estExt\_Genemark1.C\_120304 (PI 132610) we observed deletions in the promoter and in the exon 2. In order to confirm these deletions, we designed the primers pairs NRPS1F/NRPS1R and NRPS2F/NRPS2R, respectively. With primer pairs NRPS1F/NRPS1R and NRPS2F/NRPS2R, fragments of 640 bp and 671 bp were obtained, respectively. Sequences obtained were compared with those of Acv3 reference genome and no differences were observed.

In the PKS gene estExt\_Genewise1Plus.C\_120511 (PI 173482) we observed 24 SNPs. Among these, 7 were showed to produce missense mutations. Moreover, deletions in the first exon and in the middle of the exon 7 were identified (Figure 5). The position of the SNPs in the gene were compared with the positions of the predicted protein domains to visualize their possible effects. 15 SNPs affected 6 out of the 9 domains (Figure 6). To confirm these deletions, we designed primers pairs PKS1F/PKS1R and PKS2F/PKS2R, respectively. With primer pair PKS1F/PKS1R, a fragment of 437 bp was obtained. When the sequence was compared to Acv3, the sequence of the atoxigenic strain showed insertions (14 bp), and substitutions (7 transitions and 2 transversions). With the primer pair PKS2F/PKS2R, a fragment of 480 bp was obtained and showed 2 transitions and 2 transversions in the exon 7 compared with the sequence of the toxigenic strain Acv3.

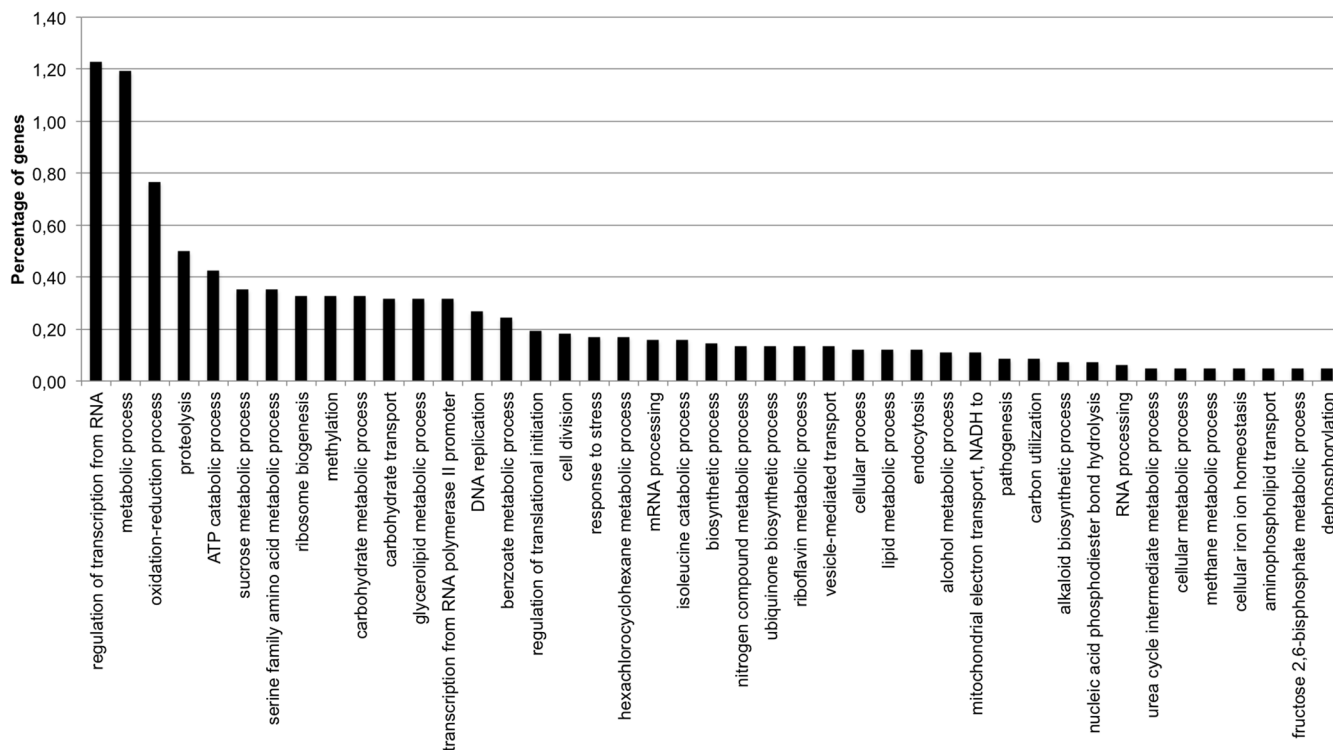
## Discussion

Despite a large amount of data is produced with the *de novo* genomes projects the real challenge of genomics, to date, is to determine the genetic differences among individuals and to understand their relationships to the phenotypic differences within species. A possible approach is to identify the relatively small changes between a genome and a reference sequence. Many sequencing technologies have been developed for resequencing projects, but few of them are suitable to study variation between individuals or species. Usually, sequencers that produce small reads with a high depth of coverage and are able to sequence with a pair-end technology are a good choice to perform resequencing studies. Unfortunately, the cost of these experiments is often not affordable by small research groups. Other cheap and accessible sequencers are available in the market and could be used for this purpose.

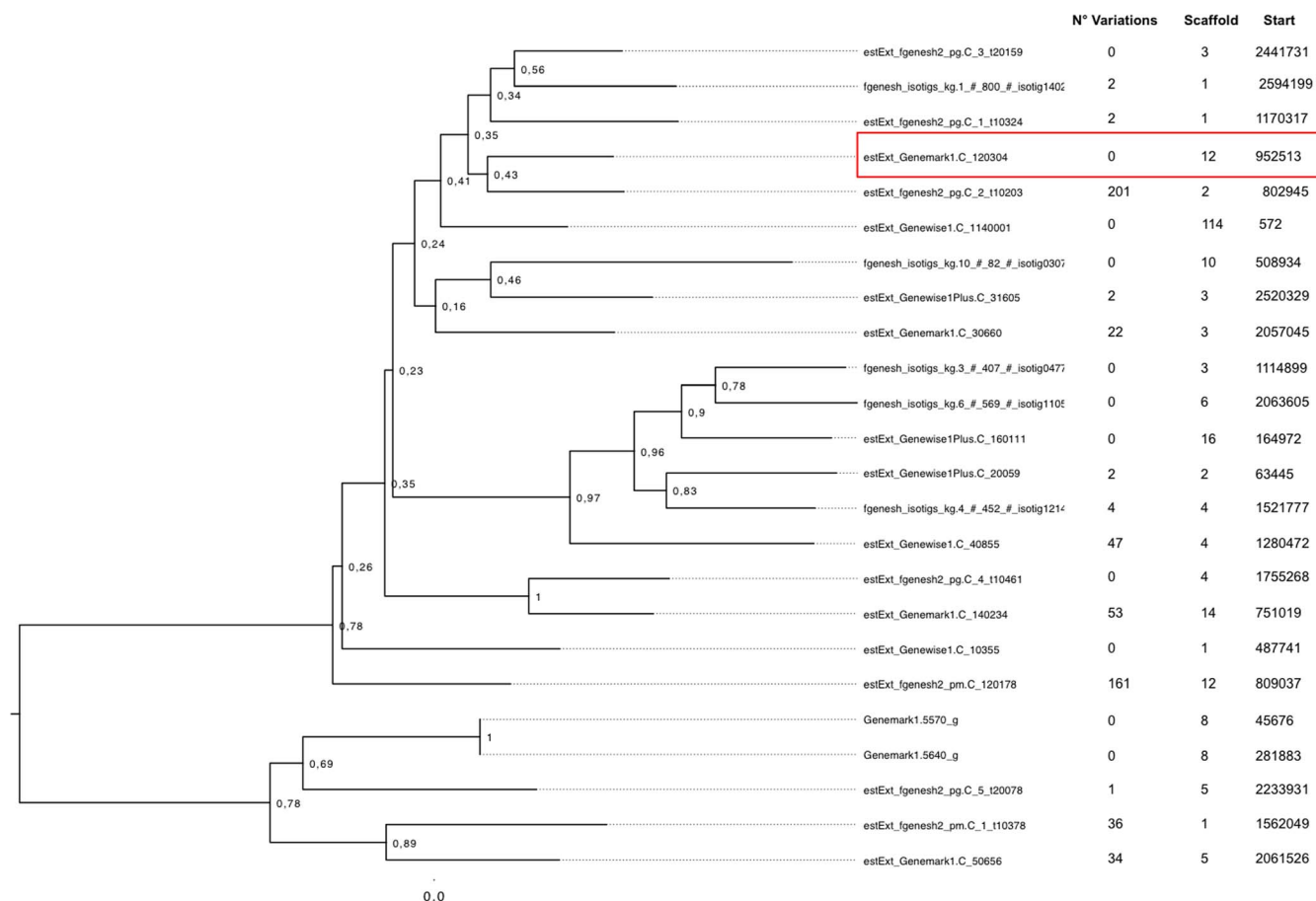
In the present study, for the first time, we used a single shotgun run of the 318 Ion Torrent chip to perform a comparison between the reference genome Acv3 and an atoxigenic wild strain of the same species. We tried to elucidate specific genes involved in OTA biosynthesis comparing the reference genome of an OTA-producing strain of *A. carbonarius* (Acv3) against a not sequenced genome of a non OTA-producing wild strain of the same species. Previous work on Acv3 revealed that the genome encodes at least 24 putative NRPS and 21 PKS encoding genes<sup>8,9</sup>. In our study a total of 24 NRPS genes and 25 PKS genes were retrieved. We also detected that in the atoxigenic strain there was a high accumulation of nonsense and missense mutations in several genes such as PKS and NRPS encoding genes. The high mutation rate of these genes could explain the lack of production of OTA by the atoxigenic strain.

**Table 2 |** SNPs and DIPs annotation

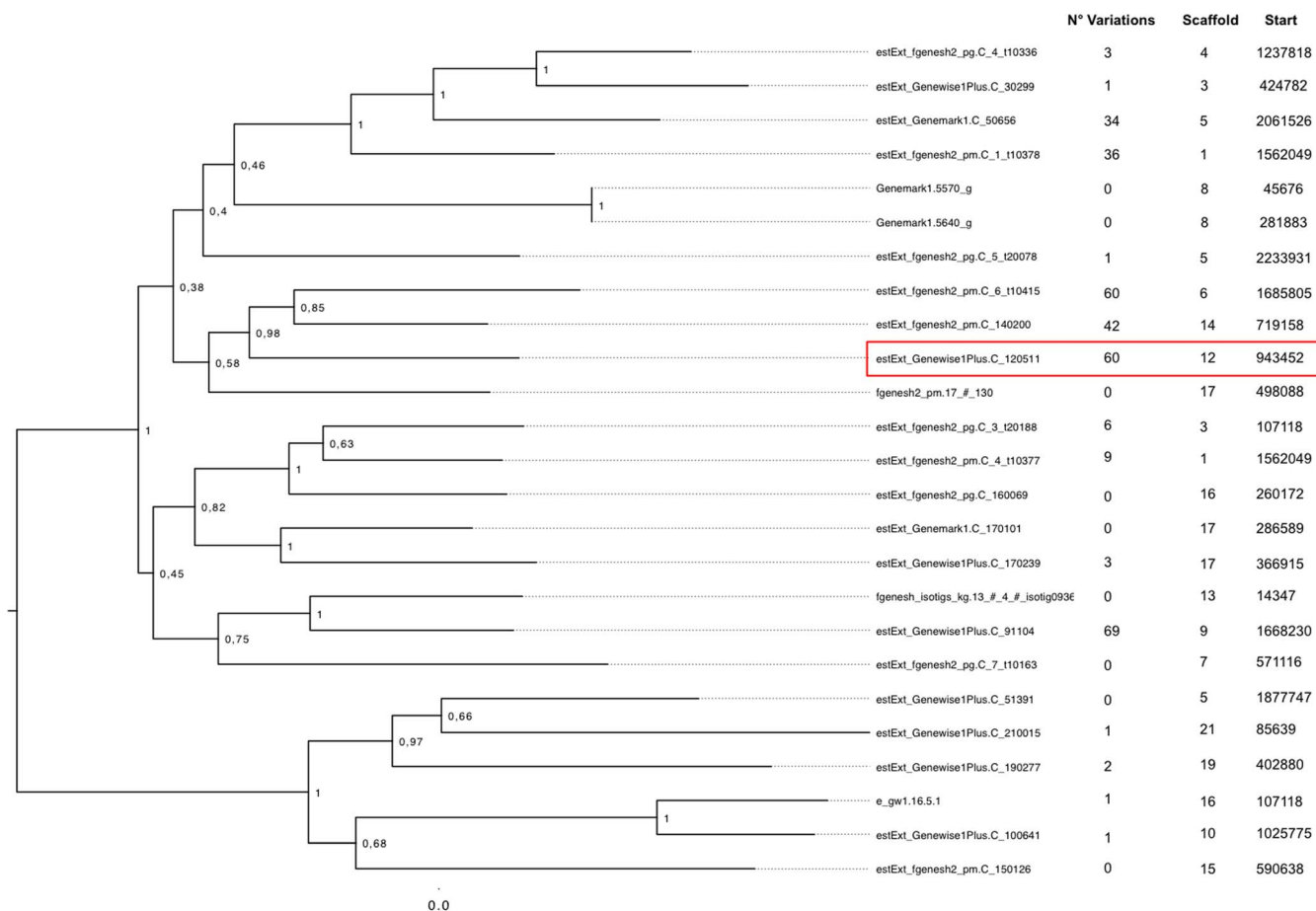
SNPs effect	N°	DIPs effect	N°
Missense	7183	Codon deletion	75
STOP codon gain	88	Codon Insertion	37
Silent	11342	Frameshift	528
Other	34048	STOP codon gain	2
Total	52661	Other	6925
Mutation Rate	1/666 bp	Mutation Rate	1/4626 bp



**Figure 2 | Gene Ontology (GO) families that have more than the 5% of mutated genes.** In the Y axis there is the percentage of gene respect the total number of genes and in the X axis the enriched families.



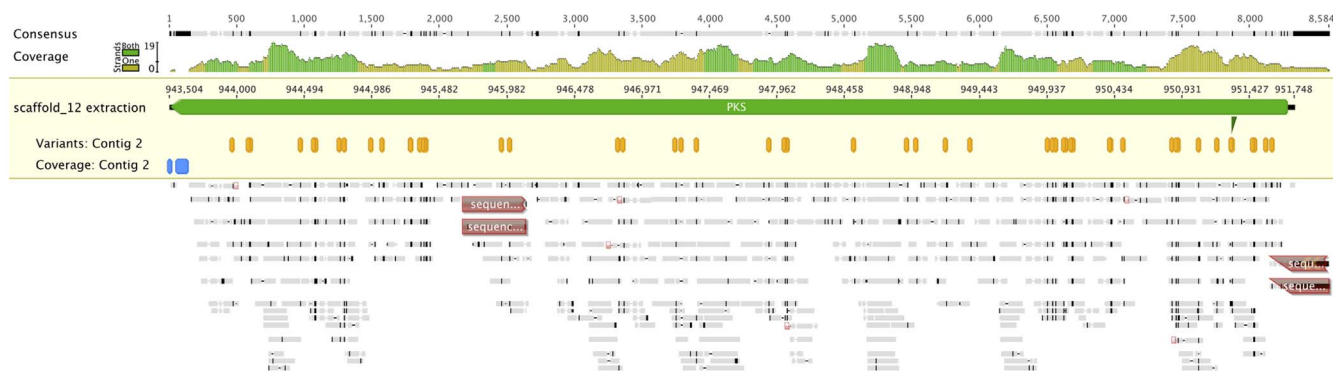
**Figure 3 | Phylogenetic reconstruction of the NRPS genes annotated on the reference genome of *A. carbonarius* (in red the NRPS estExt\_Genemark1.C\_120304 (PI 132610) gene).**



**Figure 4 | Phylogenetic reconstruction of the PKS genes annotated on the reference genome of *A. carbonarius* (in red the PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) gene).**

OTA is derived from the fungal polyketide biosynthetic pathway. Main parts of the OTA molecule are an isocoumarin moiety and the amino acid phenylalanine. However, the molecular basis of OTA biosynthesis is poorly understood. The structure of OTA suggests a biosynthetic pathway including various enzymatic steps. Besides the

biosynthesis of phenylalanine, other enzymatic activities are required in the OTA pathway such as a PKS for the synthesis of the polyketide dihydroisocoumarin, a chlorinating enzyme, a methylase, an esterase and a NRPS for ligation of the phenylalanine to the dihydroisocoumarin. So far only some OTA related PKS genes have been detected



**Figure 5 | SNPs visualization of the PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) gene.** This figure shows the PKS locus extracted from the scaffold\_12, where it is located, as a green arrow. The direction of the arrow shows that the gene is on the minus strand of the genome. The numbers above the arrow indicate the coordinates of the locus respect to the scaffold. The “Consensus” row shows the positions where no difference between the reference genome and the reads could be detected in grey, the SNPs in black and deletions as narrow grey lines. The numbers above the “Consensus” row refer to the coordinates respect to the gene. The “Coverage” row shows a histogram with the distribution of the reads along the gene body, the scale being from 0 to 19 reads. The “Variants: Contig 2” row shows the positions of the SNPs between the reference genome and the reads as yellow hexagons. The “Coverage: Contig 2” row shows the position of the deletions detected by the coverage approach as blue shapes. The bottom part of the figure shows the reads mapped on PKS. Reads portions that are equal to the reference are showed in grey, while SNPs and InDels are showed in black. Red boxes show the positions of sequencing primers.



**Figure 6** | Schematic representation of the domain composition of the PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) protein. Different domains are represented with different shapes and colors, with the relative name reported. Uncoiled regions are represented by a grey line. The identified SNPs that are affecting amino acids located in the predicted domains are reported as red marks.

in OTA producing species such as *Aspergillus ochraceus*<sup>3</sup>, *Aspergillus westerdijkiae*<sup>6</sup>, *Penicillium nordicum*<sup>4</sup>, *Penicillium verrucosum*<sup>7</sup>, *Aspergillus niger*<sup>13</sup> and *A. carbonarius*<sup>9</sup>.

Nowadays, little is known about the genes involved in the OTA biosynthesis of *A. carbonarius* which is a consistent OTA-producing species. However recently, a few wild non OTA-producing strains of *A. carbonarius* have been discovered<sup>14</sup> and some nonochratoxigenic mutant strains have been obtained both by the inactivation of genes encoding a NRPS<sup>8</sup> and a PKS<sup>9</sup> in a wild type OTA-producing strain of *A. carbonarius*, giving interesting new insights in the biosynthetic pathway of this mycotoxin. While no differences were observed in the sequences of this NRPS gene, several SNP were found in the PKS gene of the atoxigenic strain. Both genes<sup>8,9</sup> have been described to be involved in the OTA biosynthetic pathway of *A. carbonarius*. Changes detected in the PKS gene could affect the function of the protein and may be responsible for atoxigenicity in the non OTA-producing strain of *A. carbonarius*.

With our findings we demonstrate that Ion Torrent technology could be a good alternative to more expensive ones and that thanks to a specific bioinformatics pipeline it is possible to extract useful information. It seems clear that the sensitive point is to develop a specific workflow not only to manage and filter data to reach a high quality of the results but also to take advantage of single-end runs to analyze the structural variations between two genomes<sup>15,16</sup>. In this case a homemade pipeline was developed to reach high standards of variation calling using the newest published algorithms as well as specific bioinformatics methods. For the SNPs and DIPs calling we modified our pipeline, SUPERW (Simply Unified Pair-End Read Workflow), to work with single-end reads. Being not possible to use a pair-read approach, the SVs calling was performed by using a copy number variation approach, based on the differences in depth of coverage.

In order to contribute to elucidate the little-known OTA biosynthetic gene cluster of *A. carbonarius*, we plan to investigate the role of the mutations detected in PKS and NRPS encoding genes in this study. In the same direction, we plan also further metabolomic studies of both toxigenic and atoxigenic strains.

In conclusion, a deep *in silico* approach based on a comparative genomic study revealed that in the atoxigenic strain there is a high accumulation of nonsense and missense mutations in several genes. Importantly, a two fold increase in GMR was observed in PKS and NRPS encoding genes thus providing interesting clues on their role in OTA biosynthetic pathway.

## Methods

**OTA production ability detection.** OTA production was confirmed using a previously described high-pressure liquid chromatography (HPLC) screening method<sup>17</sup> designed in our laboratory. The strains A-1796 (kindly provided by P. Battilani as W-201 (=ITEM 5010) and A-2160 (=CECT 20837) from our fungal collection were first three point inoculated on Czapek Yeast extract Agar (CYA) and incubated at 15, 25 and 30 °C. After 7 and 15 days of incubation at each temperature assayed and from each strain, three agar plugs were removed from different points of the colony and extracted with 0.5 ml of methanol. The extracts were filtered and maintained at 4 °C until their analysis. Two replicates for each strain and incubation condition assayed were used. The detection limit of the extraction procedure and the HPLC technique was 0.02 ng OTA, and the limit of quantification of the HPLC technique with the extraction procedure was 0.01 µg/g for this mycotoxin.

**DNA preparation and gene sequencing analysis.** The non OTA-producing strain of *A. carbonarius* A-2160 was grown on malt extract broth medium in the dark at 25 °C for 48 h. Mycelium was recovered and grounded into fine powder using a mortar and pestle after brief nitrogen deep freezing. DNA extraction was carried out using phenol-chloroform according to the protocol described by Bragulat et al.<sup>18</sup> and treated with RNase to degrade RNA. Primers for partial amplification of NRPS estExt\_Genemark1.

C\_120304 (PI 132610) and PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) were designed with Primer3 tool<sup>19</sup>. Primers NRPS1F (5'-TTTCCCCTACTTCGTGCCAC-3') and NRPS1R (5'-CTAGAACCCCGTGCCCTTTT-3') and NRPS2F (5'-GTGGTTGTGTCTCCGGATGT-3') and NRPS2R (5'-CCGCTCCTCATGCAGAT-3') were designed to confirm the deletion in the promoter and in the exon 2 of NRPS estExt\_Genemark1.C\_120304 (PI 132610) gene, respectively. Primers PKS1F (5'-GAC-AAGGGTGGTGAGGATGG-3') and PKS1R (5'-CATTGTGCTGAGCTTTGGGC-3') and PKS2F (5'-CAACTTCCTCCTGACCGCAT-3') and PKS2R (5'-CCCACTG-CACGGATCTGAAT-3') were designed to confirm the deletion in the first exon and in the middle of exon 7 of PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) gene, respectively. The PCR products were purified with MultiScreen filter plates (Millipore, Barcelona, Spain) following the manufacturer's protocol. The purified product was used as a template for sequencing. The Big-Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and primers NRPS1F/R, NRPS2F/R, PKS1F/R and PKS2F/R were used for sequencing as specified by the manufacturer. An Applied Biosystems 3730 sequencer was employed to obtain the DNA sequences. Sequence alignments were performed using the software program Clustal X v2.0.12<sup>20</sup>.

**DNA genome Ion Torrent sequencing.** Fragment DNA library construction was performed using the Ion Xpress Plus Fragment Library Kit (Life Technologies). 100 ng of genomic DNA was used for the enzymatic shearing and ligation of Ion Torrent™ adapters according to the manufacturer's instructions. Fragments of 330 bp were selected by 2% agarose gel electrophoresis using the E-Gel® iBase™ Power System and E-Gel® Safe Imager™ (Invitrogen, Life Technologies). The quality and quantity of the genomic DNA library was assessed by analysing them in High Sensitivity DNA chip in the 2100 Bioanalyzer (Agilent). Template preparation for sequencing was made in The Ion OneTouch™ 2 System by the Ion OneTouch 200 Template Kit v2 (Life Technologies) according to manufacturer's instructions.

Sequencing was performed on the Ion Torrent Personal Genome Machine (PGM) platform using the Ion PGM 200 Sequencing Kit and an Ion 318 chip (Life Technologies).

**Whole fungal genome resequencing.** Our workflow was divided in three steps. The filtering and mapping step used as input the raw reads produced by an NGS sequencer and, utilizing the custom user parameters, automatically filters the raw reads creating a new high quality subset of reads suitable for mapping analyses<sup>21</sup>. After the filtering step, all the samples were mapped against the Acv3 reference genome using the bwa mem algorithms<sup>22</sup>. For the variation calling the mapped bam files and the reference genome are the input file of the third step of the pipeline in which the user can choose to extract small variations (SNPs and DIPs), large variations (deletions, inversions and duplications) or both<sup>23</sup>. More than 3E106 ion torrent raw reads with a maximum length of 328 bp were used in the resequencing experiment. The filter and trimming process made with an homemade pipeline based on cutadapt tool<sup>24</sup> discarded only the 0.1% of the entire reads set and decreased the reads length at 262 bp (minimum reads length 20 bp). The alignment process was conducted on the reference genome Acv3 composed by 963 scaffolds for a total size of 36 Mb. The reads were mapped on Acv3 with the bwa-mem algorithm<sup>22</sup>. About 2.8 mio of reads were mapped on the reference genome, leaving 0.3 mio of unmapped reads. About 2.3 mio mapped uniquely (Supplementary Table S4). Two alignments file were created to perform the SNPs and DIP calling and CNV analysis. The first alignment was created only with the unique mapped reads and the other with all the mapped reads for the studies. Both alignments were filtered for mapping quality and PCR-duplicates were removed. In the two mapping experiment the mean coverage is respectively 12.29× and 11.36× and the unmapped region size is of 6 Mb and 3.5 Mb highlighting that the reference genome Acv3 is composed of ~3 Mb of repetitive regions.

**SNPs and DIPs calling.** The calling was performed with samtools<sup>25</sup> on the unique mapped alignment and resulting SNPs and DIPs were filtered and annotated with SnpSift<sup>26</sup>. All the homozygous SNPs > 4 and DIPs with a depth >6 and a samtools quality > 30 were used for further analysis. The SNP and DIPs annotation and mining were performed with SNPeff<sup>27</sup> with the creation of a custom SNPeff database of Acv3 and custom parameter for the fungi genomes. As reference genes model was used the gff3 file of the reference genome Acv3.

**CNV analysis.** CNV analysis was performed with cnvator<sup>28</sup> using the multiple mapped alignment. For each scaffold the CNV regions were extracted using a bin size of 1 Kb, filtered for p-value < 0.05, duplication level <50% and divided in deleted or duplicated regions according with the coverage of the copy number varied regions (Table S5). With the coordinates of the CNVs list of duplicated or deleted genes was created.

**NRPS and PKS genes identification.** In order to identify the NRPS and PKS genes in *A. carbonarius*, the protein sequences of the PKS estExt\_Genewise1Plus.C\_120511 and of the NRPS estExt\_Genemark1.C\_120304 were used for a BLASTp search. The resulting genes were filtered by using AntiSmash and Interproscan identification of



domain composition. Only the NRPS genes showing the same domain composition of estExt\_Genemark1.C\_120304, i.e. AMP-binding (PF00501), PP-binding (PF00550), Condensation (PF00668), were used for further analyses. Similarly only the PKS genes showing the same domain composition of estExt\_Genewise1Plus.C\_120511, i.e. Ketoacyl-synt\_C (PF02801), Acyl\_transf\_1 (PF00698), PS-DH (PF14765), Methyltransf\_12 (PF08242), ADH\_N (PF08240), KR (PF08659) and PP-binding (PF00550), were used for further analyses. The identified NRPS and PKS genes are listed in the Supplementary Table S6. Domain composition of the PKS estExt\_Genewise1Plus.C\_120511 (PI 173482) protein was predicted by using SMART (<http://smart.embl-heidelberg.de/>) looking for outlier homologues and PFAM domains. The diagram of the protein and the affecting SNPs was performed with ExPasy MyDomains (<http://prosite.expasy.org/mydomains>).

**Molecular phylogenetic analysis.** Phylogeny histories of NRPS and PKS were inferred within Phylogeny.fr environment (<http://www.phylogeny.fr/> *webcite*). Each protein group was aligned with MUSCLE. Maximum Likelihood phylogeny reconstruction was employed with WAG and WAG + I + G + F models for NRPS and PKS, respectively. The bootstrap consensus trees were inferred from 100 replicates. The phylogenetic trees were visualized with FigTree 1.3.1.

**Gene Ontology analysis.** The Gene Ontology analysis has been performed with Blast2Go v 2.6.6<sup>29</sup> using the standard option and standard annotation pipeline. The results were used in REVIGO<sup>30</sup> to cleaning the GO results, create a non-redundant dataset and highlight the more representative GO categories.

- Cabañes, F. J. & Bragulat, M. R. [Ochratoxin A in profiling and speciation] *Aspergillus in the Genomic Era*. [Varga, J. & Samson, R. A. (ed.)] [57–70] (Wageningen Academic Publishers, The Netherlands, 2008).
- Ehrlich, K. C. Non-aflatoxigenic *Aspergillus flavus* to prevent aflatoxin contamination in crops: advantages and limitations. *Front. Microbiol.* **5**, 50. doi 10.3389/fmicb.2014.00050 (2014).
- O'Callaghan, J., Caddick, M. X. & Dobson, A. D. W. A polyketide synthase gene required for ochratoxin biosynthesis in *Aspergillus ochraceus*. *Microbiology* **149**, 3485–3491 (2003).
- Geisen, R., Mayer, Z., Karolewicz, A. & Färber, P. Development of a real time PCR system for detection of *Penicillium nordicum* and for monitoring ochratoxin A production in foods by targeting the ochratoxin polyketide synthase gene. *Syst. Appl. Microbiol.* **27**, 501–507 (2004).
- Schmidt-Heydt, M., Richter, W., Michulec, M., Buttinger, G. & Geisen, R. Comprehensive molecular system to study the presence, growth and ochratoxin A biosynthesis of *Penicillium verrucosum* in wheat. *Food Addit. Contam.* **25**, 989–996 (2008).
- Bacha, N., Atoui, A., Mathieu, F., Liboz, T. & Lebrihi, A. *Aspergillus westerdijkiae* polyketide synthase gene “aoks1” is involved in the biosynthesis of ochratoxin A. *Fungal Gen. Biol.* **46**, 77–84 (2009).
- Abbas, A. *et al.* Functional characterization of the polyketide synthase gene required for ochratoxin A biosynthesis in *Penicillium verrucosum*. *Int. J. Food Microbiol.* **161**, 172–181 (2013).
- Gallo, A. *et al.* New insight in the ochratoxin A biosynthetic pathway by deletion of an nrps gene in *Aspergillus carbonarius*. *Appl. Environ. Microbiol.* **78**, 8208–8218 (2012).
- Gallo, A. *et al.* Identification and characterization of the polyketide synthase involved in ochratoxin A biosynthesis in *Aspergillus carbonarius*. *Int. J. Food Microbiol.* **179**, 10–17 (2014).
- Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* **42**, D26–31 (2014).
- Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).
- Mellman, A. *et al.* Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* **6**, e22751 (2011).
- Peñ, H. J. *et al.* Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.* **25**, 221e231 (2007).
- Cabañes, F. J., Bragulat, M. R. & Castellá, G. Characterization of nonochratoxigenic strains of *Aspergillus carbonarius* from grapes. *Food Microbiol.* **36**, 135–141 (2013).
- Quail, A. M. *et al.* A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
- Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
- Bragulat, M. R., Abarca, M. L. & Cabañes, F. J. An easy screening method for fungi producing ochratoxin A in pure culture. *Int. J. Food Microbiol.* **71**, 139–144 (2001).
- Bragulat, M. R., Martínez, E., Castellá, G. & Cabañes, F. J. Ochratoxin A and citrinin producing species of the genus *Penicillium* from feedstuffs. *Int. J. Food Microbiol.* **126**, 43–48 (2008).
- Untergasser, A. *et al.* Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
- Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinformatics* **15**, 256–278 (2014).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: 1303.3997 [q-bio.GN]* (2013).
- Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–375 (2011).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Cingolani, P. *et al.* Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **3**, 35 (2012).
- Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 1–13 (2012).
- Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
- Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
- Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

## Acknowledgments

We thank P. Battilani (Università Cattolica del Sacro Cuore, Italy) for kindly providing us the strain W-201 (=A-1796 = ITEM 5010) used in this study. The sequence *Aspergillus carbonarius* ITEM 5010 v3 data used in this publication were generated by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility and is supported under Contract No. DE-AC02-05CH11231 (Principal Investigator: Scott E. Baker). This research was supported by the Ministerio de Ciencia e Innovación of the Spanish Government (AGL2010-22182-C04-02).

## Author contributions

F.J.C., M.R.B. and G.C. designed the work; G.C. performed PCR studies; M.R.B. performed mycological and HPLC studies; A.S. coordinated genome resequencing studies; W.S. and R.A.C. performed bioinformatic studies; F.J.C. coordinated manuscript writing.

## Additional information

**Accession Codes:** The genome resequencing information of *A. carbonarius* A-2160 has been deposited in the European Nucleotide Archive (ENA) under accession PRJEB6789.

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cabañes, F.J. *et al.* Rapid genome resequencing of an atoxigenic strain of *Aspergillus carbonarius*. *Sci. Rep.* **5**, 9086; DOI:10.1038/srep09086 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>