The Plantain Proteome, a Focus on Allele Specific Proteins Obtained from Plantain Fruits

Nádia A. Campos, Rony Swennen, and Sebastien C. Carpentier*

Proteomics has been applied with great potential to elucidate molecular mechanisms in plants. This is especially valid in the case of non-model crops of which their genome has not been sequenced yet, or is not well annotated. Plantains are a kind of cooking bananas that are economically very important in Africa, India, and Latin America. The aim of this work was to characterize the fruit proteome of common dessert bananas and plantains and to identify proteins that are only encoded by the plantain genome. We present the first plantain fruit proteome. All data are available via ProteomeXchange with identifier PXD005589. Using our in-house workflow, we found 37 alleles to be unique for plantain covered by 59 peptides. Although we do not have access (yet) to whole-genome sequencing data from triploid banana cultivars, we show that proteomics is an easily accessible complementary alternative to detect different allele specific SNPs/SAAPs. These unique alleles might contribute toward the differences in the metabolism between dessert bananas and plantains. This dataset will stimulate further analysis by the scientific community, boost plantain research, and facilitate plantain breeding.

Studies in plant biology through proteomics have increased considerably in the recent years. The main cause is that proteomics provides an insight into the metabolism, and is thus complementary to genomics results.\[1\] It is known that the correlation between mRNA and protein at the same moment of extraction is often low.\[2,3\] High throughput proteomics for non-model plants has been used to bypass this problem and to generate more applicable results.\[4,5\] Polyploidy and allopolyploidy considerably complicate the proteome analysis of crops. Bananas and plantains are polyploid crops originated from two wild diploid species: *Musa acuminata* (AA), which is highly polymorphic, with spindly plants that grow in clumps, and *Musa balbisiana* (BB), a more homogeneous hardy plant with a massive pseudo-trunk. There are nowadays diploid, triploid, or tetraploid genome groups.\[6,7\] The main genome groups are AA, AB, AAA, AAB, and ABB. Most dessert banana cultivars are AAB or AAA. The Cavendish subgroup, that is sold on the export market has an AAA genome constitution while plantains are AAB. Plantains are sweet acid starchy bananas with typically long fruits and are mostly consumed after frying or boiling. Plantains are an important staple crop in West and Central Africa, India, and Latin America.\[6\] Both dessert bananas and plantains are considered a non-model crop and the complexity of their genomes makes it challenging to analyze the transcriptome and the proteome.\[9\] We used here an easy and reproducible protocol for protein extraction and identification and we present the first proteome of plantain fruits (AAB). We created our own workflow to tackle the difficulties of working with a triploid non-model species without an available database. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium Partner Repository via PRIDE with the dataset identifier PXD005589. These results will stimulate further analysis by the scientific community and will boost plantain research and facilitate breeding.

Plantains fruits and Cavendish fruits were bought in the local supermarket in Leuven, Belgium. Five biological replicates (fruits) of each cultivar were selected based on their phenotypic characteristics and the same green peel color. All fruits were kept separately, cleaned, peeled, their pulp was cut into thin slices and immersed immediately in liquid nitrogen. All ten samples where lyophilized to a water content of 2.5%. After drying, the samples were hermetically sealed and stored at room temperature until the proteomics analysis was performed. Banana tissues are considered difficult for protein extraction due to the presence of many interfering compounds which makes the extraction process more difficult.\[12\] Lyophilization provides easier material for manipulation without losses in protein content and is an easy and safe way to transport the samples.\[12\] Protein extractions were performed according to the phenol extraction/ammonium acetate precipitation method we published and adapted for gel free proteomics.\[14\]
Twenty μg of proteins were digested with trypsin (Trypsin Protease, MS Grade Thermo Scientific) and purified by Pierce C18 Spin Columns (Thermo Scientific). The digested samples (0.5 μg/5 μL) were separated in an Ultimate 3000 (Thermo Scientific) UPLC system and then in a Q Exactive Orbitrap mass spectrometer (Thermo Scientific) as described.\[15\] The Q Exactive Orbitrap mass spectrometer (Thermo Scientific, USA) was operated in positive ion mode with a nano spray voltage of 1.5 kV and a source temperature of 250 °C. Proteo Mass LTQ/FT-Hybrid ESI Pos. Mode Cal Mix (MS CAL5-1EASUPELCO, Sigma-Aldrich) was used as an external calibrant and the lock mass 445.12003 as an internal calibrant. The instrument was operated in data-dependent acquisition (DDA) mode with a survey MS scan at a resolution of 70 000 (fw hm at m/z 200) for the mass range of m/z 400–1600 for precursor ions, followed by MS/MS scans of the top ten most intense peaks with +2, +3, +4, and +5 charged ions above a threshold ion count of 16 000 at 17 500 resolution using normalized collision energy (NCE) of 25 eV with an isolation window of 3.0 m/z and dynamic exclusion of 10 s. All data were acquired with Xcalibur 3.0.63 software (Thermo Scientific). For protein identification, we used MASCOT version 2.2.06 (Matrix Science) against our in house Musa A-B database containing acuminata AA proteins (dh PahangV1), the non-redundant unique balbisiana BB proteins (PKW) (http://banana-genome.hub.southgreen.fr/) and the usual contaminants for mass spectrometry (76 220 proteins). The parameters used to search were: parent mass tolerance of 10 PPM, fragment tolerance of 0.02 Da, oxidation of M as variable modification, carbamidomethyl C as fixed modification and up to one missed cleavage was allowed for trypsin. Results from MASCOT were imported to Scaffold version 3.6.5. In Scaffold, the threshold was set to minimum one peptide identified with 95% confidence and the false discovery rate (FDR) was automatically calculated based on default parameters from the software.

Using our Musa A-B database we identified in total 2144 different proteins with 0.2% FDR (Supporting Information, Table 1). Taking into account only the proteins identified in at least two biological replicates reduces this number to 1731, of which 1344 proteins were identified in Cavendish fruits and 1363 in plantain fruits (Supporting Information, Table 1). Esteve et al.\[16\] utilized the proteominer beads to identify the proteome of Cavendish fruits (Supporting Information, Table 1). The aim of our study was to characterize the proteome of plantain fruits and compare it to Cavendish fruits to identify important allele specific proteins in a cultivar that is not sequenced, plantain. The main contrasting characteristics between plantain and Cavendish are undoubtedly related to unique alleles that can explain together with epigenetic regulations the different phenotypes.\[17\] To find allele specific peptides in plantain fruits, we used a basic but very useful principle: spectral counting (Scaffold). Potential plantain allele specific peptides were filtered using the following conditions. Maximum spectral count in Cavendish = 0, which means the peptide was never identified in Cavendish; median spectral count in plantain ≠ 0, being identified at least in three biological replicates. To detect single amino acid polymorphisms (SAAPs) in acuminata (A) and balbisiana alleles (B), the identified plantain unique peptides were filtered further. Only peptide sequences that were exclusively identified in a B derived protein accession were accepted. Their allelic acuminata homolog was searched using the Greenphyl homolog function (http://www.greenphyl.org/cgi-bin/get_homologs.cgi) to determine the SAAP. Only plantain specific proteins where the acuminata homolog was successfully identified were accepted (Supporting Information, Table 2). This allowed us to allocate a protein as an A and B allele version. Further annotations of the proteins were retrieved from Uniprot software (http://www.uniprot.org/uploadlists/). Analysis of gene functions from the allelic specific proteins were made through GO enrichment annotations via our in house software (https://labtrop.shinyapps.io/UniGO/).

Following our workflow, we identified 37 interesting loci spread over all 11 chromosomes (Table 1). We appointed 59 peptides as B allele specific and 47 peptides as A allele specific. The introduction of M. balbisiana genes is said to be correlated to hardness, drought tolerance, a changed nutritional value, increased starchiness, and different maturation process.\[18–20\] Through evolution, mutations in the coding region of a gene are likely to have a different biological function, especially if the mutations occur in the protein domain, since they are generally considered as the basic units of protein folding, evolution, and function.\[24\]

Ramu et al.\[25\] highlighted some possible deleterious mutations in domesticated cassava using whole genomic screening experiments of wild ancestors and cultivars. Like banana, cassava cultivars are clonally propagated and this genomic screening study suggests that many deleterious mutations have not been crossed out. We expect a similar situation in banana. Advanced whole genomic screening experiments enable the identification and interpretation of mutations at the genome level.\[24,25\] Although we do not have access (yet) to whole-genome sequencing data from triploid banana cultivars, we show that proteomics is an easily accessible complementary alternative to detect the different allele specific SNPs/SAAPs.

To our knowledge, this is the first proteomic investigation in plantain fruits, and the most extensive fruit proteomic study in the genus Musa. This public release of the plantain fruit proteome is an important step for plantain varietal selection and breeding.
| Uniprot entry and protein annotation | BB accession number | AA accession number | B allele b,c | A allele b,c,d |
|--------------------------------------|---------------------|---------------------|--------------|--------------|
| M0S0K5 Protein disulfide-isomerase | KMMuBchr1_C01477    | GSMUA_Achr1T16970_001 | AASLSKNDPPVVLAK | AASLSKNDPPVVLAK |
|                                      |                     |                     | EADGWEYLLK    | EADGWEYLLK    |
|                                      |                     |                     | LHEVAENYGGK   | LHEVAENYGGK   |
| M0S1P5 Uncharacterized protein      | KMMuBchr1_C01832    | GSMUA_Achr1T20870_001 | AGVENMFCSVGGIPVTAVATR | AGVENMFCSVGGIPVTAVATR |
|                                      |                     |                     | ATRHTIQPR     | No ID         |
|                                      |                     |                     | GYLAGTPTEELKSALSEFSAR | GYLAGTPTEELKSALSEFSAR |
|                                      |                     |                     | LNLWHLFGQPPR  | No ID         |
|                                      |                     |                     | SHPWEAISKK    | No ID         |
| M0RFLU7 D-3-phosphoglycerate dehydrogenase | KMMuBchr1L_C29495 | GSMUA_Achr1T07062_001 | CGLGMHVISHPYPAPDR | CGLGMHVISHPYPAPDR |
| M0RKL2 Uncharacterized protein      | KMMuBchr1Q_C30906   | GSMUA_Achr1T07287_001 | LVLPGLAK      | No ID         |
| M0RPm3 Pectinesterase               | KMMuBchr1L_C32183   | GSMUA_Achr1T07543_001 | SNTNLMMFDGIGK | No ID         |
| M0RQ6L6 Uncharacterized protein     | KMMuBchr1L_C25307   | GSMUA_Achr1T08680_001 | IVQDQSVLQDEKR | IVQDQSVLQDEKR |
| M0RS60 Uncharacterized protein      | KMMuBchr1L_C33444   | GSMUA_Achr1T11430_001 | YGVPDATLDILNTIAR | YGVPDATLDILNTIAR |
| M0RT8B Uncharacterized protein      | KMMuBchr1L_C33678   | GSMUA_Achr1T11830_001 | KIEDELSSSHEK  | No ID         |
| M0RTW1 Uncharacterized protein      | KMMuBchr1L_C33846   | GSMUA_Achr1T20310_001 | LVPVGYCIC     | LVPVGYCIC     |
|                                      |                     |                     | TYISGDSQSKDDVR | TYISGDSQSKDDVR |
|                                      |                     |                     | WYDSVQGLAFRPCK | WYDSVQGLAFRPCK |
| M0RNAJ0 Ubiquitin carboxyl-terminalhydrolase | KMMuBchr1L_C34086 | GSMUA_Achr1T23000_001 | FWEESFLDFRLFYK | FWEESFLDFRLFYK |
| M0SSQ5 Uncharacterized protein      | KMMuBchr2G_C36699   | GSMUA_Achr2T06640_001 | SYITQIQSKDDAVYALSALATSPSADYVINVAR | SYITQIQSKDDAVYALSALATSPSADYVINVAR |
| M0SSR0 Uncharacterized protein      | KMMuBchr2G_C3706    | GSMUA_Achr2T06980_001 | LDSAESMLKDFLNK | LDSAESMLKDFLNK |
| M0SA45 4-alpha-gluconotransferase   | KMMuBchr2G_C30504   | GSMUA_Achr2T22060_001 | TGDLPDVPYDFTRPSVDPTR | TGDLPDVPYDFTRPSVDPTR |
| M0SAT8 Formate dehydrogenase, mitochondrial | KMMuBchr3G_C05265 | GSMUA_Achr3T01200_001 | AAEAGLTVAYTCGSNVVSVAEDELMR | AAEAGLTVAYTCGSNVVSVAEDELMR |
|                                      |                     |                     | LKPFNCLNLNYDR  | LKPFNCLNLNYDR  |
| MN0588 Succrose synthase             | KMMuBchr3G_C05330   | GSMUA_Achr3T01900_001 | SVPLAAGDCAAFNSAK | SVPLAAGDCAAFNSAK |
|                                      |                     |                     | VVHGIDVDPFKNVSPSDLTITYFFYTEK | VVHGIDVDPFKNVSPSDLTITYFFYTEK |
| M0SC42 Pectinesterase               | KMMuBchr3G_C05670   | GSMUA_Achr3T07540_001 | LPRPGQINTITACGR | LPRPGQINTITACGR |
| M0SCZ5 Uncharacterized protein      | KMMuBchr3G_C05965   | GSMUA_Achr3T07870_001 | SYPVNRTNASSEK  | SYPVNRTNASSEK  |
|                                      |                     |                     | TIKDMVLSSER   | No ID         |
| M0SCK8 Uncharacterized protein      | KMMuBchr3G_C07974   | GSMUA_Achr3T27620_001 | AVTLELKK      | No ID         |
| M0SRS5 Malate dehydrogenase         | KMMuBchr4G_C10674   | GSMUA_Achr4T21920_001 | NAIWGNHSSTQYPDCHATVK | NAIWGNHSSTQYPDCHATVK |
| M0SRMS Uncharacterized protein      | KMMuBchr4G_C10876   | GSMUA_Achr4T24140_001 | AFDSYEAVLKDPDGDAVYPLPTSLHLR | AFDSYEAVLKDPDGDAVYPLPTSLHLR |
|                                      |                     |                     | AIGALAPNIVWAGSR | AIGALAPNIVWAGSR |
|                                      |                     |                     | HLLLEKPTLCAALDLR | HLLLEKPTLCAALDLR |
|                                      |                     |                     | WAAWAAECKG    | WAAWAAECKG    |
| M0SVOB Uncharacterized protein      | KMMuBchr5G_C12124   | GSMUA_Achr5T05290_001 | GYYIQPTIFDSVEDKMK | GYYIQPTIFDSVEDKMK |
| M0WS53 Protein transport Sec61 subunit beta | KMMuBchr5G_C12409 | GSMUA_Achr5T08400_001 | ARGSSQSQTTASAGGARPAVPR | ARGSSQSQTTASAGGARPAVPR |

(Continued)
Table 1. Continued.

| Uniprot entry and protein annotation | BB accession number | AA accession number | B allele | A allele |
|--------------------------------------|---------------------|--------------------|----------|----------|
| M05ZK9 Nucleoside diphosphate kinase | KMMuBchr5_G13683    | GSMUA_Achr5T18300,001 | GLVGEIINRFK | GLVGEIHSR |
| M0TIH5 Uncharacterized protein      | KMMuBchr5_G14355    | GSMUA_Achr5T25000,001 | NVIGSDSIEGASK | NVIGSDSIEGAR |
| M05HD1 Importin subunit alpha       | KMMuBchr5_G14553    | GSMUA_Achr3T24150,001 | HVTITAFSK | HVTITAFSK |
| M0TA69 Uncharacterized protein      | KMMuBchr6_G17518    | GSMUA_Achr6T25730,001 | SPPIEVIQAGWPVR | No ID |
| M0TB62 Methylthioribose-1-phosphate isomerase | KMMuBchr6_G17823 | GSMUA_Achr6T29170,001 | EASEKHHHLF | EAEASGKH HHHLF |
| M0TCA9 Uncharacterized protein      | KMMuBchr6_G18191    | GSMUA_Achr7T01530,001 | AALADQDAEFFSLAE | ALADQKEFFSANAAQASR |
| M0TDU3 Uncharacterized protein      | KMMuBchr7_G18687    | GSMUA_Achr7T01530,001 | CVKPPVIFDVSRPK | CVKPIHYGVDSRPK |
| M0TEY6 Uncharacterized protein      | KMMuBchr7_G19042    | GSMUA_Achr7T05460,001 | EGQYGAGICPGYDIHSPR | No ID |
| M0TX4 Uncharacterized protein       | KMMuBchr7_G20737    | GSMUA_Achr7T19370,001 | IATIUSDLVATLDAEQCKELK | IATILYSDVLATLDAEQCKELK |
| M0JC2 Uncharacterized protein       | KMMuBchr7_G20876    | GSMUA_Achr7T20850,001 | RLGVLGCSSTSWAAYR | RLGVLGCSSTSWAAYR |
| M0TRR4 Uncharacterized protein      | KMMuBchr8_G23675    | GSMUA_Achr8T19110,001 | QFNSIPICGEMKEDVQPK | QFNSIPICGEMKEDVQPK |
| M0T866 Uncharacterized protein      | KMMuBchr8_G23861    | GSMUA_Achr8T20830,001 | IMYADEADPFCAASDGCGDRNMILGR | IMYADEADPFCAASDGCGDRNMILGR |
| M0U2C0 Uncharacterized protein      | KMMuBchr9_G27509    | GSMUA_Achr9T20200,001 | MKEIAEYLGSYVK | MKEIAEYLGSYVK |
| M0U78 Uncharacterized protein       | KMMuBchrUn_random_G35868 | GSMUA_AchrUn_random | GLLSCTGTVGVSMFANKFPR | GLLSCTGTVGVSMFANKFPR |
| M0UCP7 Uncharacterized protein      | KMMuBchrUn_random_G39488 | GSMUA_AchrUn_random | NLKEDHDELCADIVQK | No ID |

a) Chr in the name of the accessions refers to the chromosome number of the locus.
b) All peptides have been identified with a probability > 99%.
c) SAAPs are indicated in bold.
d) No ID: the allelic variant peptide was not confidently identified in our experiment.
Abbreviations

DDA, data dependent acquisition; EST, expressed sequence tag; FDR, false discovery rate; GO, gene ontology; MS, mass spectrometry; NCE, normalized collision energy; PPM, parts per million; SAAP, single amino acid polymorphism; SNP, single nucleotide polymorphism

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors would like to thank Kusay Arat for the technical support at SYBIOMA, KU Leuven, Belgium. We also thank Annick de Troyer and Madelyn Ibana for operating the hyophiler and Kostas Tsolis for lab support. We acknowledge USAID for the project AID-BFS-C-II-00002-11 Reviving the plantain breeding program at International Institute for Tropical Agriculture (IIITA). The authors would furthermore like to thank all donors who supported this work through their contributions to the CGIAR Fund (http://www.cgiar.org/who-we-are/cgiar-fund/fund-donors-2/), and in particular to the CGIAR Research Program on Roots, Tubers, and Bananas.

Conflict of Interest

The authors have declared no conflict of interest.

Keywords

allele, amino acid polymorphisms, non-model proteomics, plantain

Received: June 12, 2017
Revised: December 31, 2017

[1] J. M. Palma, F. J. Corpora, L. A. del Río, J. Proteomics 2011, 74, 1230.
[2] R. C. Janssen, J.-P. Nap, L. Mlynarova, Nat. Biotechnol. 2002, 20, 19.
[3] S. C. Carpentier, B. Coemans, N. Podevin, K. Laukens, E. Witter, H. Matsumura, R. Terauchi, R. Swennen, B. Panis, Physiol. Plant. 2008, 133, 117.
[4] S. C. Carpentier, T. America, Plant Proteomics, 2014, 1072, 333.
[5] M. Zivy, S. Wienkoop, J. Renaut, C. Pinheiro, E. Goulas, S. Carpentier, Front. Plant Sci. 2015, 6, 448.
[6] R. Ortiz, R. Swennen, Biotechnol. Adv. 2014, 32, 158.
[7] S. B. Janssens, F. Vandelook, E. De Langhe, B. Verstraete, E. Smets, I. Vandenhouwe, R. Swennen, New Phytol. 2016, 210, 1453.
[8] G. Aurore, B. Parfait, L. Fahrasmane, Trends Food Sci. Technol. 2009, 20, 78.
[9] Y. Zorrilla-Fontanesi, M. Rouard, A. Cenci, E. Kissel, Y. Do, E. Dubois, S. Nidelet, N. Roux, R. Swennen, S. C. Carpentier, Sci. Rep. 2016, 6, 1.
[10] J. A. Vizcaíno, E. W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, J. A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P. A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R. J. Chalkley, H.-J. Krauss, J. P. Albar, S. Martinez-Bartolome, R. Apweiler, G. S. Oommen, L. Martens, A. R. Jones, H. Hermjakob, Nat. Biotechnol. 2014, 32, 223.
[11] J. A. Vizcaíno, R. C. Gote, A. Csordas, J. A. Dianes, A. Fabregat, J. M. Foster, J. Griss, E. Alpi, M. Birirn. J. Contell, G. O’Kelly, A. Schoenegger, D. Ovelleiro, Y. Pérez-Riverol, F. Reisinger, D. Rios, R. Wang, H. Hermjakob, Nucleic Acids Res. 2013, 41, D1063.
[12] S. C. Carpentier, K. Dens, I. Van den houwe, R. Swennen, B. Panis, Proteomics 2007, 7 Suppl 1, 64.
[13] S. C. Carpentier, E. Witters, K. Laukens, P. Deckers, R. Swennen, B. Panis, Proteomics 2005, 5, 2497.
[14] K. Buts, S. Michielssens, M. L. A. T. Hertog, E. Hayakawa, J. Cordewener, A. H. P. America, B. M. Nicolai, S. C. Carpentier, J. Proteomics 2014, 105, 31.
[15] N. A. Campos, L. V. Paiva, B. Panis, S. C. Carpentier, Proteomics 2016, 16, 1001.
[16] C. Esteve, A. D’Amato, M. L. Marina, M. C. García, P. G. Righetti, Electrophoresis 2013, 34, 207.
[17] I. Hippolyte, C. Jenny, L. Gardes, F. Bakry, V. Pomies, P. Cubry, K. Tornekpe, A. M. Risterucci, N. Roux, M. Rouard, E. Arnaud, M. Kolesnikova-Allen, X. Perrier, Ann. Bot. 2012, 109, 937.
[18] N. W. Simmonds, K. Shepherd, Bot. J. Linn. Soc. 1955, 55, 302.
[19] J. C. Robinson, V. C. G. Saúco, Bananas and plantains, 2nd ed., CABI, Oxfordshire 2010.
[20] S. R. Choudhury, S. Roy, P. P. Saha, S. K. Singh, D. N. Sengupta, Plant Cell Rep. 2008, 27, 1235.
[21] A.-C. Vanhove, W. Vermaelen, A. Cenci, R. Swennen, S. C. Carpentier, Data Br. 2015, 3, 78.
[22] P. Kumar, S. Henikoff, P. C. Ng, Nat. Protoc. 2009, 4, 1073.
[23] M. D. Swanson, D. M. Boudreaux, L. Salmon, J. Chugh, H. C. Winter, J. L. Meagher, S. André, P. V. Murphy, S. Oscarson, R. Roy, S. King, M. H. Chapman, I. J. Goldstein, E. B. Tarbet, B. L. Hurst, D. F. Snee, C. de la Fuente, H.-H. Hoffmann, Y. Xue, C. M. Rice, D. Schols, J. V. Garcia, J. A. Stuckey, H.-J. Gabius, H. M. Al-Hashimi, D. M. Markovitz, Cell 2015, 163, 746.
[24] A. Heger, L. Holm, J. Mol. Biol. 2003, 328, 749.
[25] P. Ramu, W. Esuma, R. Kawuki, I. Y. Rabbi, C. Egesi, J. V. Breeden, S. B. Janssens, F. Vandelook, E. De Langhe, B. Verstraete, E. Smets, I. Vandenhouwe, R. Swennen, New Phytol. 2016, 210, 1453.