

New criteria for the molecular identification of cereal grains associated with archaeological artefacts

Andre Carlo Colonese¹, Jessica Hendy^{1,2}, Alexandre Lucquin¹, Camilla Speller¹, Matthew Collins^{1,3}, Francesco Carrer⁴, Regula Gubler⁵, Marlu Kühn⁶, Roman Fischer⁷, Oliver E. Craig¹

1. BioArCh, Department of Archaeology, University of York, York, YO10 5DD, United Kingdom.
2. Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, 07745, Germany
3. EvoGenomics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark
4. McCord Centre for Landscape, School of History, Classics and Archaeology, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom
5. Archaeological Service of the Canton of Bern, CH-3001, Bern, Switzerland
6. Integrative Prähistorische und Naturwissenschaftliche Archäologie (IPNA), Spalenring 145, CH-4055, Basel, Switzerland
7. Target Discovery Institute, University of Oxford, NDMRB, Oxford, OX3 7FZ

Supplementary information

Sample	Extraction	Lipid detected	Attribution
137337	AE	FA (C ₁₂ -C ₃₂ , C _{16:1} -C _{18:1} , C _{17br} -C _{20br}), ALK (C ₁₁ , C ₂₀ -C ₂₉), HFA (C ₁₆ , C ₂₀ , 22-25), Stigmastane-3,5-diene, campesterol, ergostanol, β-sitosterol, stigmastanol, 10-oxo-C _{18:0}	Plant
137338	AE	FA (C ₉ -C ₂₆ , C _{16:1} -C _{20:1} , C _{14br} -C _{17br}), ALK (C ₁₁ , C ₂₃), HFA (C ₂₄), Stigmastane-3,5-diene, campesterol, β-sitosterol, stigmastanol, DT, DC (C ₈₋₁₁ , C ₂₀)	Plant (conifer)
137339	AE	FA (C ₁₀ -C ₂₄ , C _{16:1,2} -C _{20:1} , C _{15br} -C _{17br}), ALK (C ₁₁), Stigmastane-3,5-diene, β-sitosterol, DT, DC (C ₈₋₉ , C ₂₀)	Plant (conifer)
137337	TLE	FA (C ₁₄ -C ₂₂), ALK (C ₂₀ -C ₂₉), Stigmastadiene, Stigmastane-3,5-diene, campesterol, stigmastanol, β-sitosterol, stigmast-4-en-3-one, AR (C ₁₇ -C ₂₅)	Cereal
137338	TLE	FA (C ₁₄ -C ₂₂ , C _{18:1} , C _{17br}), ALK (C ₂₀ -C ₂₄), Stigmastane-3,5-diene, β-sitosterol, DT	Plant (conifer)
137339	TLE	FA (C ₁₄ -C ₂₂ , C _{18:1} , C _{17br}), ALK (C ₂₃), Stigmastane-3,5-diene, β-sitosterol, DT	Plant (conifer)

SI1. Samples selected for GCMS and details of the absorbed lipids detected. FA (Cx:y) - fatty acids with carbon length x and number of unsaturations y, br -branched chain acids; DC - α,ω-dicarboxylic acids; ALK (Cx) - n-alkanes with carbon length x; HFA – Hydroxy fatty acids; 10-oxo-C_{18:0} - 10-oxo-octadecanoic acid; DT, Diterpenoid derivatives - isopimaric acid, abietic acid, 7-oxodehydroabietic acid and dehydroabietic acid. AR(Cx) - alkylresorcinols with carbon length x.

Protein	Species	Common name	Function	Mascot Score	Peptide	BLAST Results (Lowest Common Ancestor)
Basic endochitinase C and Basic endochitinase A*	<i>Hordeinae</i>	Barley/Rye	Defence	283	R.YCDILR.V	Not unique to plants
					R.AIGVDLLR.N	Not unique to plants
					Y.CFKQER.G + Deamidated (NQ)	Not unique to plants
					R.AIGVDLLR.N.P	Mesangiospermae
					R.AIGVDLLR.N.P + Deamidated (NQ)	Mesangiospermae
					P.GFGTTGSTDTR.K	Matches to <i>S. cereale</i> , <i>H. bulbosum</i> & <i>T. aestivum</i>
					R.AAGRVPFGFVITN.I	Pentapetalae
					N.GGIECGHGQDSR.V	Pooideae
					P.DLVATDPTVSFK.T	Mesangiospermae
					A.GAFPGFGTTGSTDTR.K	Matches <i>S. cereale</i> & <i>T. aestivum</i>
					H.AVITGQWSPSGTDR.A + Deamidated (NQ)	Triticeae
					R.NPDLVATDPTVSFK.T	Mesangiospermae
					N.IVNGGIECGHGQDSR.V	Pooideae
					N.IVNGGIECGHGQDSR.V	Pooideae
					N.IVNGGIECGHGQDSR.V + Deamidated (NQ)	Pooideae
					N.IINGGLECGHGQDSR.V	Pooideae
					N.IINGGLECGHGQDSR.V	Pooideae
					N.IINGGLECGHGQDSR.V + Deamidated (NQ)	Pooideae
					T.NIVNGGIECGHGQDSR.V	Pooideae
					F.VAAAGAFPGFGTTGSTDTR.K	Matches <i>S. cereale</i> & <i>T. aestivum</i>
					D.LLRNPDLVATDPTVSFK.T	Hordeinae
					G.VGYGDNLDCYNQRPFA.-	Mesangiospermae
					G.VGYGDNLDCYNQRPFA.- + Deamidated (NQ)	Mesangiospermae
					R.VGYGNNLDCYNQRPFA.- + 2 Deamidated (NQ)	Mesangiospermae

					L.GVGYGDNLDCYNQRPFA.-	Mesangiospermae
					R.GATSNYCTPSAQWPCAPGK.S + Deamidated (NQ)	Hordeinae
					R.GATSNYCTPSAQWPCAPGK.S + 2 Deamidated (NQ)	Hordeinae
					R.AIGVDLLRNPDLVATDPTVSFK.T	Hordeinae
					R.AIGVDLLRNPDLVATDPTVSFK.T	Hordeinae
					R.AIGVDLLRNPDLVATDPTVSFK.T + Deamidated (NQ)	Hordeinae
					R.AIGVDLLRNPDLVATDPTVSFKTA + Deamidated (NQ)	Hordeinae
					R.YCDILGVGYGDNLDCYNQRPFA.-	Triticeae
					R.YCDILGVGYGDNLDCYNQRPFA.- + Deamidated (NQ)	Triticeae
					R.YCDILGVGYGDNLDCYNQRPFA.- + 2 Deamidated (NQ)	Triticeae
					R.YCDILGVGYGDNLDCYNQRPFA.- + Deamidated (NQ)	Triticeae
					A.AGRVPGFGVITNIVNGGIECGHGQDSR.V + 2 Deamidated (NQ); Propionamide (N-term)	Triticeae
Probable aquaporin TIP3-2	BOP clade	Grasses	Facilitates water transport	169	N.ILAGGPFDDGAAMNPAR.A + Oxidation (M)	BOP clade
					A.NILAGGPFDDGAAMNPAR.A	Poaceae
					A.NILAGGPFDDGAAMNPAR.A + Oxidation (M)	Poaceae
					A.GANILAGGPFDDGAAMNPAR.A + Oxidation (M)	Poaceae
Serpins Z1 A/B/C*	<i>Triticum aestivum</i>	Wheat	Defence	154	R.LSIAHQTR.F	Not unique to plants
					R.LASTISSNPK.S	<i>Triticum/Aegilops</i>
					Q.GTEAAASTAIK.M	<i>Triticum/Aegilops</i>
					K.LSAEPDFLER.H	<i>Triticum/Aegilops</i>
					R.EDISGVVLFMGMH.V	<i>Triticum/Aegilops</i>
					R.EDISGVVLFMGMH.V + Oxidation (M)	<i>Triticum/Aegilops</i>
					K.YKADTQSVDFQTK.A + Deamidated (NQ)	<i>Triticum aestivum</i>
					F.VEVNEQGTEAAASTAIK.M	<i>Triticum/Aegilops</i>
					F.MSSMDDQYLSSSDGLK.V + Oxidation (M)	Matches <i>T. urartu</i> & <i>T.</i>

						<i>aestivum</i>
					A.FVEVNEQGTEAAASTAIK.M + Propionamide (N-term)	Matches <i>T. urartu</i> & <i>T. aestivum</i>
					Q.AFVEVNEQGTEAAASTAIK.M + Deamidated (NQ)	<i>Triticum/Aegilops</i>
					Q.AFVEVNEQGTEAAASTAIK.M	<i>Triticum/Aegilops</i>
					R.VSSVFHQAFVEVNEQGTEAAASTAIK.M	<i>Triticum/Aegilops</i>
Serpine Z2*	<i>Triticum</i>	Wheat	Defence	87	R.LSIAHQTR.F	Not unique to plants
					R.LVLGNALYFK.G	Poaceae
					R.LVLGNALYFK.G + Deamidated (NQ)	Poaceae
					R.EDTSGVVLFIGH.V	<i>T. aestivum, urartu & Aegilops</i>
					A.FVEVNETGTEAAATTIAK.V + Propionamide (N-term)	<i>T. aestivum, urartu & Aegilops</i>
					K.AFVEVNETGTEAAATTIAK.V + Deamidated (NQ)	<i>T. aestivum, urartu & Aegilops</i>
Alpha-amylase/trypsin inhibitor CM3*	<i>Triticum</i>	Wheat	Defence	134	R.EMQWDFVR.L	<i>T. aestivum & durum</i>
					R.YFIALPVPSQPVD.P	<i>T. aestivum/Aegilops & durum</i>
					N.VGESGLIDLPGCPR.E	<i>T. aestivum/Aegilops & durum</i>
					R.LLVAPGQCNLATIH.N	<i>T. aestivum/Aegilops & durum</i>
					R.YFIALPVPSQPVDPR.S	<i>T. aestivum/Aegilops & durum</i>
					R.YFIALPVPSQPVDPR.S + Deamidated (NQ)	<i>T. aestivum/Aegilops & durum</i>
					R.SGNVGESGLIDLPGCPR.E	<i>T. aestivum/Aegilops & durum</i>
					R.SGNVGESGLIDLPGCPR.E + Deamidated (NQ)	<i>T. aestivum/Aegilops & durum</i>
					K.LYCCQELAEISQQR.C	<i>T. aestivum & durum</i>
Alpha-amylase inhibitor 0.53*	<i>Triticeae</i>	Wheat	Defence	119	K.LQCNGSQVPEAVLR.D	Triticeae
					K.LQCNGSQVPEAVLR.D + Deamidated (NQ)	Triticeae
					K.LQCNGSQVPEAVLR.D + 2 Deamidated (NQ)	Triticeae
					R.LPIVVDASGDGAYVCK.D	Triticeae
Probable aquaporin PIP2-1	BOP clade	Grasses	Facilitate	114	Y.TVFSATDPKR.N	Embryophyta

			water transport		A.TIPITGTGINPAR.S	Embryophyta
					R.YGGGANTLAAGYSK.G	BOP clade
					R.YGGGANTLAAGYSK.G + Deamidated (NQ)	BOP clade
					H.LATIPITGTGINPAR.S	Embryophyta
Xylanase inhibitor protein 1	<i>Triticum/Aegilops</i>	Wheat	Defence	94	R.SALDLFDHL.W	Triticeae
					K.TGQVTVFWGR.N	Poaceae
					R.CGYPAAHVGR.A	Triticeae
					R.CGYPAAHVGR.A	Triticeae
					K.NVYYGVAPVAQK.K	<i>Triticum/Aegilops</i>
					K.DNYGGIMLWDR.Y	Triticeae
					K.NVYYGVAPVAQKK.D	<i>Triticum/Aegilops</i>
					L.SIGGYGTGYSLSNR.S	Triticeae
Wheatwin 1/2*	<i>T. aestivum, urartu & Aegilops</i>	Wheat	Defence	93	N.PATGAQITAR.I	Triticeae
					N.PATGAQITAR.I + Propionamide (N-term)	Triticeae
					R.VTNPATGAQITAR.I	Triticeae
					R.LSIAHQTR.F	Not unique to plants
					R.LVLGNALYFK.G	Poaceae
					R.LVLGNALYFK.G + Deamidated (NQ)	Poaceae
					R.EDTSGVVLFIGH.V	<i>T. aestivum, urartu & Aegilops</i>
					A.FVEVNETGTEAAATTIAK.V + Propionamide (N-term)	<i>T. aestivum, urartu & Aegilops</i>
					K.AFVEVNETGTEAAATTIAK.V + Deamidated (NQ)	<i>T. aestivum, urartu & Aegilops</i>
ADP, ATP carrier protein, mitochondrial	Poaceae	Grasses	ADP/ATP carrier	78	R.GNTANVIR.Y	Not unique to plants
					R.GNTANVIR.Y + Deamidated (NQ)	Not unique to plants
					R.AIAGAGVLSGYDQLQIL.F	Poaceae
					A.LGWLITNGAGLASYPIDTVR.R + Deamidated (NQ); Propionamide (N-term)	Mesangiospermae
Agglutinin isolectin 1 / Agglutinin isolectin 3*	<i>Triticum</i>	Wheat	Carbohydrate binding	75	R.VCTNNYCCSK.W	Matches <i>T. urartu, durum & aestivum</i>
					K.WGSCGIGPGYCGAGCQSGGCD.G	Triticeae
					K.WGSCGIGPGYCGAGCQSGGCDG.V	Triticeae
					K.WGSCGIGPGYCGAGCQSGGCDG.V + Deamidated (NQ)	Triticeae

Glucan endo-1,3-beta-glucosidase GI*	<i>Triticeae</i>	Wheat	Defence	75	R.FDAVTNT.F + Deamidated (NQ)	Not unique to plants
					R.VVVSSESGWPSASGFAATADNAR.A	Triticeae
Gamma-gliadin B*	<i>Triticeae</i>	Wheat	Major seed storage protein	75	G.IIQPQQPAQLEVIR.S	Triticeae
					Q.GIIQPQQPAQLEVIR.S	Triticeae
					Q.GIIQPQQPAQLEVIR.S + Deamidated (NQ)	Triticeae
Chymotrypsin inhibitor WCI*	<i>Triticum/Aegilops</i>	Wheat	Defence	70	R.ELAAISSNCR.C + Deamidated (NQ)	<i>Triticum/Aegilops</i>
					R.TLALPGQCNLPTIHGGPY.C	<i>Triticum/Aegilops</i>
					A.FPPSQSQGGGPPQPPLAPR.C + Deamidated (NQ); Propionamide (N-term)	<i>Triticum/Aegilops</i>
Trypsin/alpha-amylase inhibitor CMX1/CMX2/CMX3*	<i>T. aestivum, urartu & Aegilops</i>	Wheat	Defence	64	R.EFIAGIVGR.E	Not unique to plants
					R.QTCGYLSAER.Q	<i>T. aestivum, urartu & Aegilops</i>
					R.QTCGYLSAER.Q + Deamidated (NQ)	<i>T. aestivum, urartu & Aegilops</i>
Alpha-amylase/trypsin inhibitor CM16*	<i>Triticeae</i>	Wheat	Defence	58	R.EVQMDFVR.I	Triticeae
					R.DYVEQQACR.I + Deamidated (NQ)	Triticeae
					K.SRPDQSGLMELPGCPR.E	Triticeae
					K.QQCCGELANIPQQCR.C	Triticeae
					K.QQCCGELANIPQQCR.C + Deamidated (NQ); Propionamide (N-term)	Triticeae
					R.ILVTPGYCNLTTHNTPYCLAMEE.S + Oxidation (M)	<i>T. aestivum, durum, macha</i>
Oleosin*	Poaceae	Grasses	Stabilizes lipids	63	R.QAFQRT.P	Not unique to plants
					R.TPDYVEEAR.R	Poaceae
					R.QAFQRTPDY.V	Poaceae
					R.QAFQRTPDYVEE.A + Deamidated (NQ)	Poaceae
					R.QAFQRTPDYVEE.A + 2 Deamidated (NQ)	Poaceae
					R.QAFQRTPDYVEEA.R	Poaceae
					R.QAFQRTPDYVEEA.R + 2 Deamidated (NQ)	Poaceae
					R.QAFQRTPDYVEEA.R + Deamidated (NQ)	Poaceae
					R.QAFQRTPDYVEEAR.R	Poaceae
					R.QAFQRTPDYVEEAR.R	Poaceae
Non-specific lipid-transfer protein 2G and 2P*	<i>Triticum</i>	Wheat	Lipid transfer	63	K.DPTYGQYIR.S	<i>T. aestivum/Aegilops/durum</i>
					K.PSGECCGNLR.A	Triticeae

					R.AQQGCFCQYAK.D	BOP clade
					R.DTLTSCGLAVPHC.-	<i>T. aestivum</i> & <i>durum</i>
					R.DTLQSCGLAVPHC.- + Deamidated (NQ)	<i>T. aestivum</i> & <i>durum</i>
					S.AILSGAKPSGECCGNLR.A	<i>T. aestivum/Aegilops/durum</i>
					S.AILSGAKPSGECCGNLR.A	<i>T. aestivum/Aegilops/durum</i>
Germin-like protein 3-3*	BOP clade	Grasses	Defence	62	D.AGVVELLK.S	Not unique to plants
					R.VDAGVVVELLK.S	BOP clade
					K.ALRV D A G V V E L L K . S	BOP clade
					K.ALRV D A G V V E L L K . S	BOP clade
Alpha-amylase inhibitor 0.28*	<i>Triticeae</i>	Wheat	Defence	56	R.SVYQELGVR.E	Triticeae
					R.CGDLSSMLR.S	Triticeae
					R.CGDLSSMLR.S + Oxidation (M)	Triticeae
Oxalate oxidase*	<i>Triticeae</i>	Wheat	Cell wall structure	54	E.AGVVELLK.S	Not unique to plants
					L.TLFGSDPPIPTPVLT.K.A	<i>Hordeum vulgare</i> **
Puroindoline-B*	<i>Triticeae</i>	Wheat	Defence	51	K.CNMGADCKFPGS.Y	Pooideae
					K.LSSCKDYVMER.C + Oxidation (M)	Triticeae
					R.AQSLPSKCNMGADCKFPGS.Y + Oxidation (M); Deamidated (NQ)	Pooideae
Gamma-gliadin*	<i>Triticinae</i>	Wheat	Major seed storage protein	51	L.VSILPR.S	Not unique to plants
					S.LVSILPR.S	Not unique to plants
					S.SLVSILPR.S	Mesangiospermae
					F.LQQMNPK.N	Triticeae
					Q.GLGIIQPQPAQLEGIR.S	Triticinae
Serine carboxypeptidase 2*	<i>Triticum</i>	Wheat	Enzyme	47	Y.ITVDEGAGR.S	<i>T. urartu</i> & <i>aestivum</i>
					Y.GASEELGAFR.V	Poales
					R.SLFYLLQEAPEDAQ.P	<i>Triticum/Aegilops</i>
					R.LPGQPAVDFDMYSGYITVDEGAGR.S + Oxidation (M)	<i>T. urartu</i> & <i>aestivum</i>
					R.LPGQPAVDFDMYSGYITVDEGAGR.S + Oxidation (M); Deamidated (NQ)	<i>T. urartu</i> & <i>aestivum</i>
Cysteine proteinase inhibitor 8 / Cysteine proteinase inhibitor 6*	Mesangiospermae	Flowering plants	Enzyme	47	S.GEQQVVS GMNYR.L	Mesangiospermae
					S.GEQQVVS GMNYR.L + Oxidation (M)	Mesangiospermae
Aspartic proteinase oryzasin-1	BOP clade	Grasses	Breakdown	43	N.YMNAQYFGEIGVGTPPQK.F + Oxidation (M); 2	Petrosaviidae

or Phytapsin*	n of storage proteins	Deamidated (NQ)	
		K.NYMNAQYFGEIGVGTPPQK.F + Oxidation (M);	Petrosaviidae
		Deamidated (NQ)	
		K.VGEGAAAQCISGFTAMDIPPPR.G + Oxidation (M)	BOP clade

SI2. Plant proteins identified in the residue (#137337). Mascot scores are an indication of protein identification significance and are generated based on the probability of a random match. Scores are stated as $-10 \cdot \log_{10}(P)$, where P is the absolute probability. *Proteins found expressed in the grain of the plant. ** Deamidation of N and D are not able to be to be unambiguously distinguished, so an assignment to *Hordeum vulgare* cannot be confidentially determined.