

# Supporting Information

Jeandet et al. 10.1073/pnas.1500783112

## SI Materials and Methods

**Champagne Sampling.** Six different wines were used for this set of experiments: three old champagnes discovered in Baltic Sea (A11, A33, and B17) and three modern champagnes from 1980 (1980), 1955 (1955), and 2011 (BCJ) from VCP House. A11 and A33 were two VCP champagnes, with the latter contaminated by seawater. B17 was a champagne from the producer Juglar. Data concerning those champagnes (Table S1) were obtained according to current enological analyses by VCP.

**Elemental Analysis.** For the analysis, 250- $\mu$ L aliquots of the samples were introduced into quartz vessels, and 1 mL of suprapure, subboiling distilled HNO<sub>3</sub> (Merck) was added. The vessels were closed and introduced into a pressure digestion system (Seif) for 10 h at 170 °C. The resulting clear solution was diluted to exactly 5 mL with Milli-Q H<sub>2</sub>O and was then ready for element determination. Since only a small sample volume was available, the resulting 1:20 dilution of the samples (250 /5,000  $\mu$ L) provided enough sample volume for measurements and the digestion eliminated matrix effects and sensitivity changes due to ethanol from samples.

**ICP-AES.** An ICP-AES SPECTRO CIROS Vision system (SPECTRO Analytical Instruments GmbH & Co. KG) was used for a complete elemental screening. The measured spectral lines were: Al, 167.078 nm; As, 189.042 nm; B, 249.773 nm; Ba, 455.404 nm; Bi, 223.061 nm; Ca, 396.847 nm; Cd, 214.438 nm; Co, 228.616 nm; Cr, 267.716 nm; Cu, 324.754 nm; Fe, 259.941 nm; Hg, 184.950 nm; K, 766.490 nm; Mg, 279.553 nm; Mn, 257.611 nm; Mo, 202.030 nm; Na, 589.592 nm; Ni, 231.604 nm; P, 177.495 nm; Pb, 220.353 nm; S, 180.731 nm; Sb, 206.833 nm; Se, 196.090 nm; Sn, 189.991 nm; Sr, 407.771 nm; V, 292.464 nm; W, 207.922 nm; and Zn, 213.856 nm. Samples were introduced using a peristaltic pump equipped with an antipulse head (SPETEC), connected to a Meinhard nebulizer with a cyclon spray chamber.

The RF power was set to 1,000 W, and the plasma gas flow rate was 15 L Ar/min, whereas the nebulizer gas flow rate was 0.6 L Ar/min.

**ICP-sf-MS.** In cases where element analysis by ICP-AES resulted in readings close to or below the detection limit, samples were reanalyzed by an ELEMENT 2, ICP-sf-MS instrument (Thermo). Analyzed element isotopes were <sup>111</sup>Cd, <sup>208</sup>Pb, <sup>202</sup>Hg, <sup>95</sup>Mo, <sup>121</sup>Sb, <sup>182</sup>W in low-resolution mode, <sup>59</sup>Co, <sup>52</sup>Cr, <sup>63</sup>Cu, <sup>56</sup>Fe, <sup>55</sup>Mn, <sup>60</sup>Ni, <sup>64</sup>Zn in medium-resolution mode, <sup>75</sup>As, <sup>77</sup>Se, and <sup>78</sup>Se in high-resolution mode. Samples were introduced using a peristaltic pump equipped with an antipulse head (SPETEC), connected to a Meinhard nebulizer with a cyclon spray chamber. The RF power was set to 1,200 W, and the plasma gas flow rate was 15 L Ar/min, whereas the nebulizer gas flow rate was 0.9 L Ar/min.

**Quality Control for ICP-AES and ICP-sf-MS.** Every 10 measurements, three blank samples and a control sample of certified standards (SPEX CertiPrep) for all mentioned elements were run. Results were calculated on a computerized laboratory data management system, relating the sample measurements to calibration curves, blank determinations, control standards, and the weight of the digested sample.

**FTICR/MS.** High-resolution mass spectra (1) were acquired on a Bruker (Bruker Daltonics GmbH) solariX FTICR/MS equipped with a 12-Tesla superconducting magnet (Magnex Scientific Inc.) and an APOLO II electrospray ionization (ESI) source (Bruker

Daltonics GmbH) in negative and positive ionization modes. Spectra were first externally calibrated using arginine clusters (10 mg/L in methanol), and the accuracy was attained. Spectra were acquired with a time domain of four megawords and a mass range of  $m/z$  100–1,000. Five hundred scans were accumulated in negative mode and 300 scans in positive mode for each sample. FTICR spectra were internally recalibrated to fatty acids with mass errors below 0.05 ppm and exported to peak lists at signal-to-noise ratio (S/N) of 4 and higher. After calibration and peak alignment, the  $m/z$  values were annotated with unambiguous elemental formulas by in-house written software based on exact mass difference of 0.1 ppm and considering the elements C, H, N, O, P, and S and their corresponding isotopes as described in ref. 2. It is noteworthy for readers not familiar with ultrahigh mass spectrometry that this exceptional mass accuracy and mass resolution allows assignment of the exact mass of molecular ions and the respective elemental composition directly out of mixtures with precision greater than the mass of an electron (which amounts to  $9.10938215 \times 10^{-31}$  kg, or 1/1836.2 of the mass of a proton). Accordingly, two molecular compositions with mass differences smaller than that of a single electron mass can be differentiated, and elementary compositions can be conclusively assigned to their isotopologues. As many as thousands of such compositions containing C, H, O, N, and S elements can be calculated and then be represented using 2D van Krevelen diagrams, which sort them onto two axes according for instance to H/C and O/C atomic ratios (Fig. S5).

**RP and HILIC-UPLC-Q-ToF-MS.** Metabolites were separated using a Waters Acquity UPLC system coupled to a Bruker maXis UHR-ToF-MS. The RP method was used (1). Detection was carried out in negative ionization mode. Bruker raw data files were imported into Genedata Expressionist for MS 7.6 and noise subtracted, aligned, and peak picked using three workflows to split workflow.

**Hierarchical Clustering Analysis.** Hierarchical clustering analysis (HCA) was performed on the normalized data using the Hierarchical Clustering Explorer 3.5 software. This method was able to group the samples into homogeneous and distinct clusters without imposing preliminary hypotheses on the data. Principal Component Analysis (PCA) also disclosed the principal groups of data. PCA is another unsupervised method with the capacity to reduce the complexity of a multivariate dataset. The masses discriminative for the two groups were calculated with an Orthogonal PLS model (OPLS). PCA and OPLS were performed with the SIMCA 13.0.3 software (Umetrics) for FTICR/MS data. Genedata Expressionist has been applied for MS Analyst 7.6 for LC-MS data. The 2D van Krevelen diagrams were constructed using compositional networks (based on elemental compositions) and functional networks based on selected functional group equivalents, enabling improved assignment option of elemental composition and classification of organic complexity with tunable validation windows.

**Network Analysis of FTICR/MS Data.** The matrix that resulted from the six unified mass spectra in ESI(-)-FTICR/MS was filtered for missingness (features with less than two nonzero intensities were excluded), which resulted in a matrix of 11,242  $m/z$  signals. Mass defect filtering and correlation based deisotoping excluded 3,269 features. Network reconstruction was performed on the remaining 7,973 features, first with an edge error of 0.1 ppm and then with an edge error of 0.2 ppm (100 ppb and 200 ppb, respectively). Before network-based feature annotation as presented in ref. 2, all given

$m/z$  signals were converted into the carbon, hydrogen, nitrogen, oxygen, phosphorus, sulfur space using the following parameters: C100, O80, N10, S3, P3, and  $\pm 0.5$  ppm error, generating 12,915 sum formulas after filtering the Senior rules. Consequent isotopic pattern matching enabled the definition of six very high confidence formulas for the initiation of network annotation. Then 4,196  $m/z$  features were annotated using 0.1-ppm edge error (Fig. 3).

**CSN.** The given formulas were deconvoluted into vectors that contained all element/C and element/H ratios for each formula (multiplex van Krevelen vectors). On the basis of these vectors, it was possible to construct CSNs.

The CSN was colored for nonsignificant features (green), specific modern champagne features (blue), and specific old champagne features (red). The network contains 4,196 nodes ( $m/z$  features) and 20,880 edges with a modularity of 0.91 (ranging from  $-1$  and  $1$ ). This modularity score is high in comparison with other values in the literature (3, 4). This network comprises the framework for all network visualizations in Fig. 3 and Fig. S4.

**Compound Class Assignment.** The neutral formulas of the dataset were matched against compound databases. Only 8% of the features could be matched with KNAPSACK, which is in the range of usual annotations in our mass translator into pathways Kyoto Encyclopedia of Genes and Genomes-based database and which made a colored network diagram for annotation obsolete. To gain insights into possible ontological compound classes, another database was used [MeSH ([www.ncbi.nlm.nih.gov/mesh](http://www.ncbi.nlm.nih.gov/mesh))/IDEOM database (5)] including 18,159 formulas with known compound class distribution (72 classes). Each champagne formula was attributed to the most abundant compound class assignment of its compositionally most similar MeSH/IDEOM entry, and the node coloring was performed according to these compound classes. It can be seen in Fig. S4 that the compound class labels (colors) form a multitude of modules (clusters).

**Compound Class Enrichment Analysis for Modern and Baltic Champagnes.** Since it is tedious to present all interesting areas of the CSN, a compound class enrichment analysis for markers for modern and Baltic champagnes was performed. The enrichment analysis is performed assuming a discontinuous hypergeometric distribution, which is tested with the Fisher exact test.

**Sample Preparation for NMR Analysis.** An aliquot of 120  $\mu\text{L}$  of each champagne was mixed with 60  $\mu\text{L}$   $\text{D}_2\text{O}$  buffer [90%  $\text{D}_2\text{O}$ , 500 mM  $\text{PO}_4$ , 0.1% trimethylsilyl-tetradeuteropropionic acid (TSP), pH 7.4] and transferred to 3-mm-outer-diameter NMR Bruker Match tubes (Hilgenberg GmbH).

**NMR Analysis.** NMR spectra were acquired on a Bruker 800-MHz spectrometer (Bruker Biospin) operating at 800.35 MHz with a quadruple inverse cryoprobe. For an overview of molecules present in the sample, a standard 1D pulse sequence [recycle delay (RD)- $90^\circ$ - $t_1$ - $90^\circ$ - $tm$ - $90^\circ$ -acquire free induction decay (FID)] was acquired, with water suppression irradiation during RD of 2 s, mixing time ( $tm$ ) set on 200 ms, and a  $90^\circ$  pulse set to 8  $\mu\text{s}$ , collecting 1,024 scans into 64,000 data points with a spectral width of 12 ppm. A representative sample of one Baltic Sea champagne and one modern champagne underwent a series of 2D analyses [ $J$ -resolved (JRES), total correlation spectroscopy (TOCSY), and heteronuclear single quantum coherence (HSQC)] for detailed metabolite analysis and metabolite identification. The 1H NMR spectra

were acquired with the pulse sequence [ $d1$ - $90^\circ$ - $\tau$ - $180^\circ$ - $\tau$ -acquire FID], with suppression of the water resonance during  $d1$  (2 s) into 4,000 data points in F2 and 16 transients using 64 increments of  $\tau$ ; the spectral widths in F2 and F1 were 12 ppm and 0.08 ppm, respectively. For the 2D 1H-13C HSQC spectra, phase-sensitive  $g$ -2D-HSQC using preservation of equivalent pathways and adiabatic pulses for inversion and refocusing with gradients were used. For each 2D spectrum,  $4,096 \times 2,048$  data points were collected using 32 scans per increment, an acquisition time of 0.25 s, and 16 dummy scans. The spectral widths were set to 12 ppm and 230 ppm in the proton and carbon dimensions, respectively. For 2D 1H-1H TOCSY spectra, phase-sensitive sensitivity-improved 2D TOCSY with water suppression by gradient tailored excitation (3-9-19) and using DIPSI-2 were acquired. For each 2D spectrum,  $19,228 \times 1,024$  data points were collected using 32 scans per increment, an acquisition time of 1 s, and 16 dummy scans. The spectral widths were set to 12 ppm in the F2 and F1 dimensions. Processing of the spectra was performed using TopSpin 3.2 (Bruker BioSpin). FIDs were multiplied by an exponential decaying function corresponding to a line broadening of 0.3 Hz before Fourier transformation. All spectra were manually phased, baseline corrected, and calibrated to TSP ( $\delta$  0.00). Data were imported to Matlab (Mathworks) and normalized to the TSP signal (area under the curve of TSP).

**Quantification of Selected Metabolites.** Each champagne sample was prepared and its spectra acquired under identical laboratory and experimental conditions. The spectra were normalized to the internal standard TSP. The area under the curve of the NMR peak from metabolites of interest was calculated in Matlab and stoichiometrically corrected by dividing by the number of protons that give rise to the peak. The obtained value is an arbitrary unit.

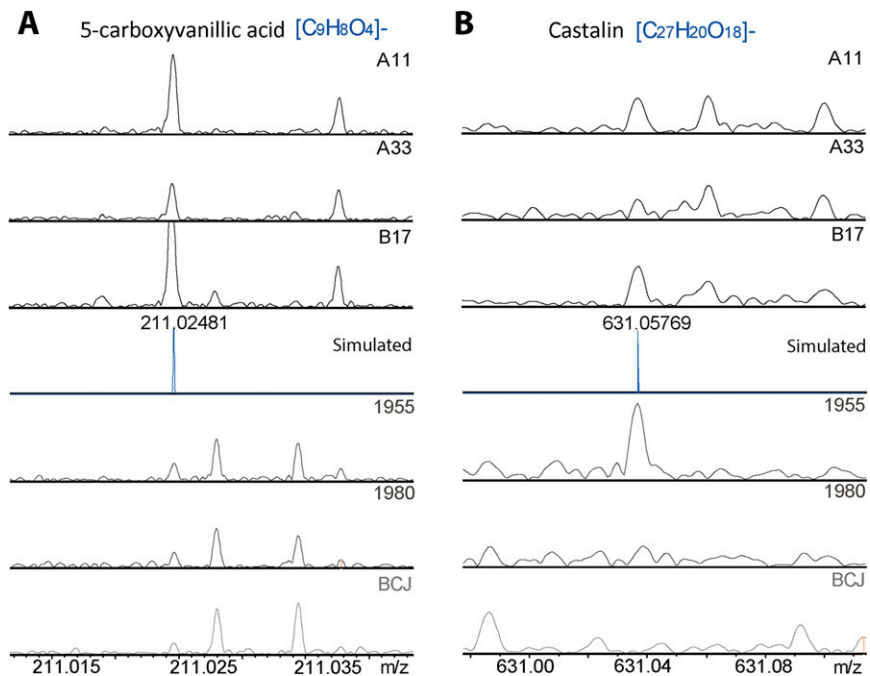
**Aroma Analyses Through SBSE-LD-GC-MS.** Volatile compounds were extracted by SBSE-LD (6). A stir bar (0.5 mm thick; 10 mm long) coated with a polydimethylsiloxane film (Twister; Gerstel) was immersed in  $3 \times 5$  mL of a Baltic Sea champagne sample and stirred for 180 min at 1,250 rpm at room temperature. For reverse extraction purposes, the stir bars were rinsed with deionized water and gently dried, then placed into 250- $\mu\text{L}$  glass flat-bottom inserts filled with 100  $\mu\text{L}$  of acetonitrile inside a glass vial. The reverse extraction was performed by ultrasonic treatment for 30 min. Volatile compounds were analyzed with 1  $\mu\text{L}$  of the extracted material in an Agilent 6890N gas chromatograph equipped with an Agilent 7683 automatic liquid sampler coupled to an Agilent 5975B inert mass selective detector (MSD) (Agilent Technologies). The gas chromatograph was fitted with a DB-Wax capillary column (60 m  $\times$  0.32 mm i.d.  $\times$  0.50  $\mu\text{m}$  film thickness; J & W Scientific), and helium was used as carrier gas (1 mL/min constant flow). The GC oven temperature was programmed without initial hold time at a rate of 2.7  $^\circ\text{C}/\text{min}$  from 70  $^\circ\text{C}$  to 235  $^\circ\text{C}$  (hold 10 min). The injector was set to 250  $^\circ\text{C}$  and used in pulsed splitless mode (25 psi for 0.50 min). The temperatures of the interface, MS ion source, and quadrupole were 270  $^\circ\text{C}$ , 230  $^\circ\text{C}$ , and 150  $^\circ\text{C}$ , respectively. The mass spectrometer was operated in electron impact ionization mode (70 eV), and the masses were scanned over an  $m/z$  range of 29–300 atomic mass units. Agilent MSD chemStation software (G1701DA, Rev D.03.00) was used for instrument control and data processing. All compounds were identified by mass spectrometry data, their linear retention index (KOVATS indexes) (7–9), and/or comparison with a standard (Table S5).

1. Roullier-Gall C, Witting M, Gougeon RD, Schmitt-Kopplin P (2014) High precision mass measurements for wine metabolomics. *Front Chem* 2:102.
2. Tziotis D, Hertkorn N, Schmitt-Kopplin P (2011) Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: Improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur J Mass Spectrom (Chichester, Eng)* 17(4):415–421.

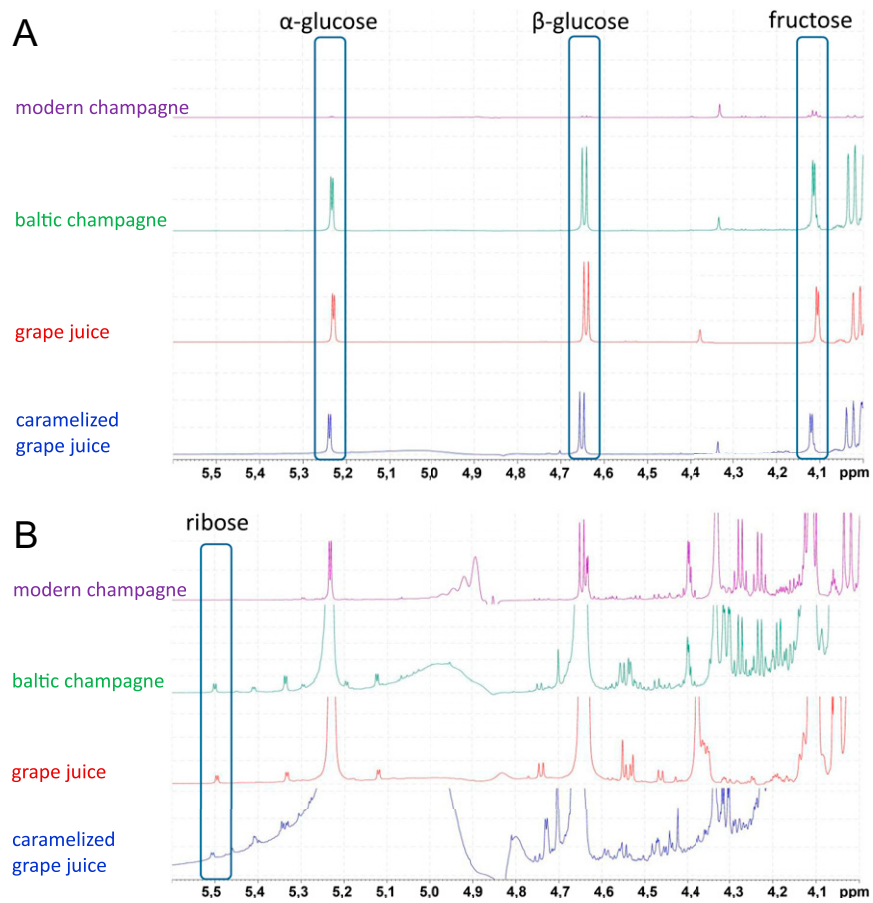
3. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103(23):8577–8582.
4. Olesen JM, Bascompte J, Dupont YL, Jordano P (2007) The modularity of pollination networks. *Proc Natl Acad Sci USA* 104(50):19891–19896.
5. Creek DJ, Jankevics A, Burgess KE, Breittling R, Barrett MP (2012) IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics* 28(7):1048–1049.

6. Steyer D, et al. (2012) QTL mapping of the production of wine aroma compounds by yeast. *BMC Genomics* 13:573.
7. Culleré L (2010) Characterisation of aroma active compounds in black truffles (*Tuber melanosporum*) and summer truffles (*Tuber aestivum*) by gas chromatography-olfactometry. *Food Chem* 122(1):300–306.

8. Pino JA, Tolle S, Gök R, Winterhalter P (2012) Characterisation of odour-active compounds in aged rum. *Food Chem* 132(3):1436–1444.
9. Csoka M, Amtmann M, Sardy DN, Kallay M, Korany K (2013) GC-MS description of the primary aroma structure of two Kadarka wines considered indigenous in Hungary. *J Appl Bot Food Qual* 86:104–112.

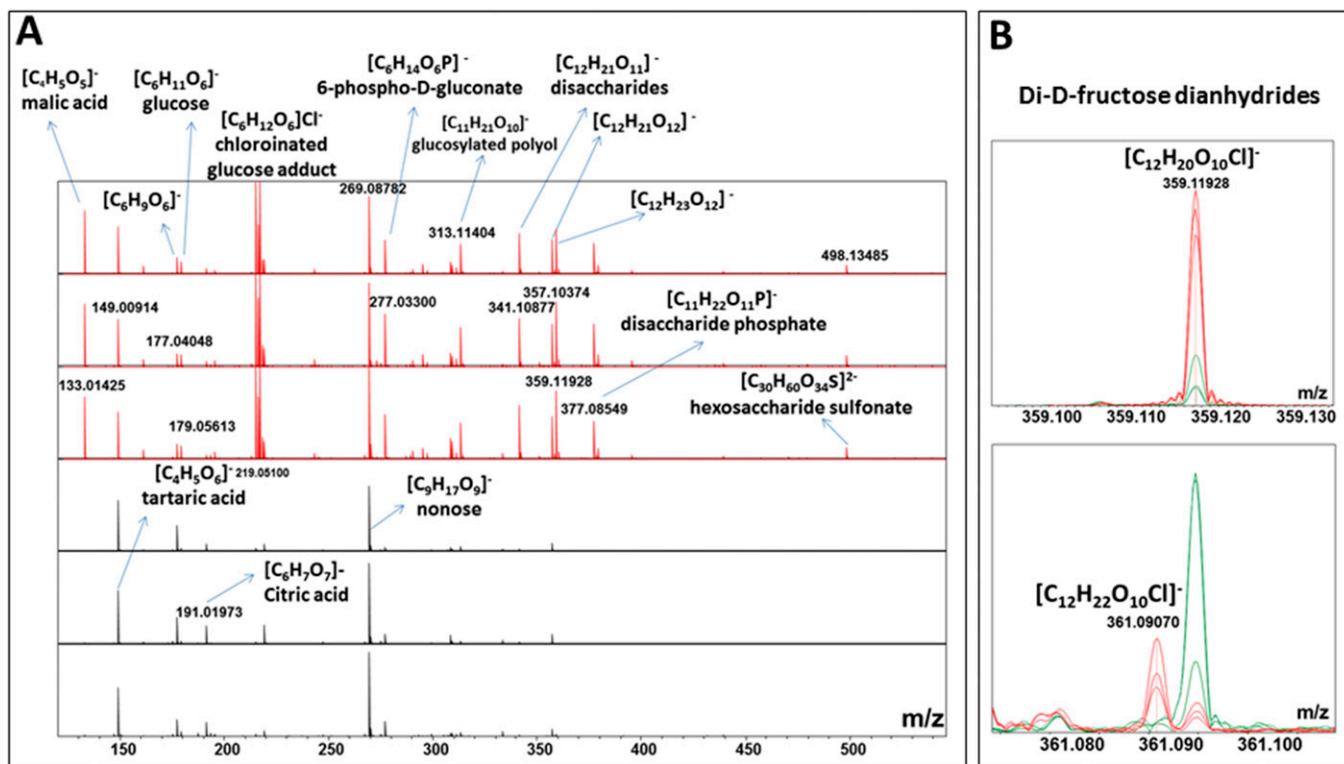


**Fig. S1.** FTICR/MS characterization of wood markers (5-carboxyvanillic acid and castalin). (A) Details of the (–) FTICR/MS mass spectra  $m/z$  211 showing the presence of the  $m/z$  211.02481 signal in all champagnes, assigned to  $[C_9H_8O_4]^-$ . (B) The  $m/z$  631 showing the presence of the  $m/z$  631.05769 signal in Baltic Sea and 1955 champagnes, assigned to  $[C_{27}H_{20}O_{18}]^-$ .

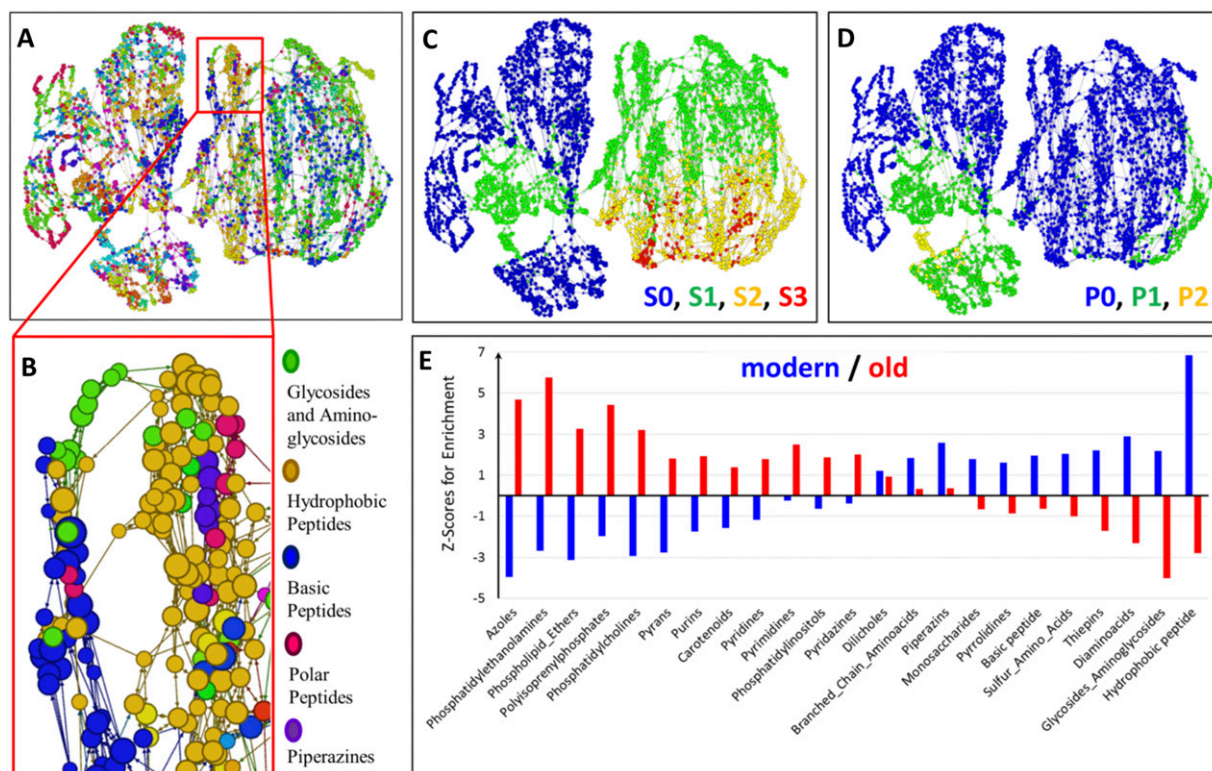


**Fig. S2.** The  $^1\text{H}$  NMR spectroscopy-generated sugar profiles. (A) Profiles including  $\alpha$ -glucose,  $\beta$ -glucose, and fructose in one representative sample each of modern champagne, Baltic Sea champagne, grape juice, and caramelized grape juice. (B) Profiles including ribose in one representative sample each of modern champagne, Baltic Sea champagne, grape juice, and caramelized grape juice. All sugars are present in Baltic Sea champagnes and both grape juices. Modern champagnes only contain very low amounts of glucose and fructose, and ribose was not detected.



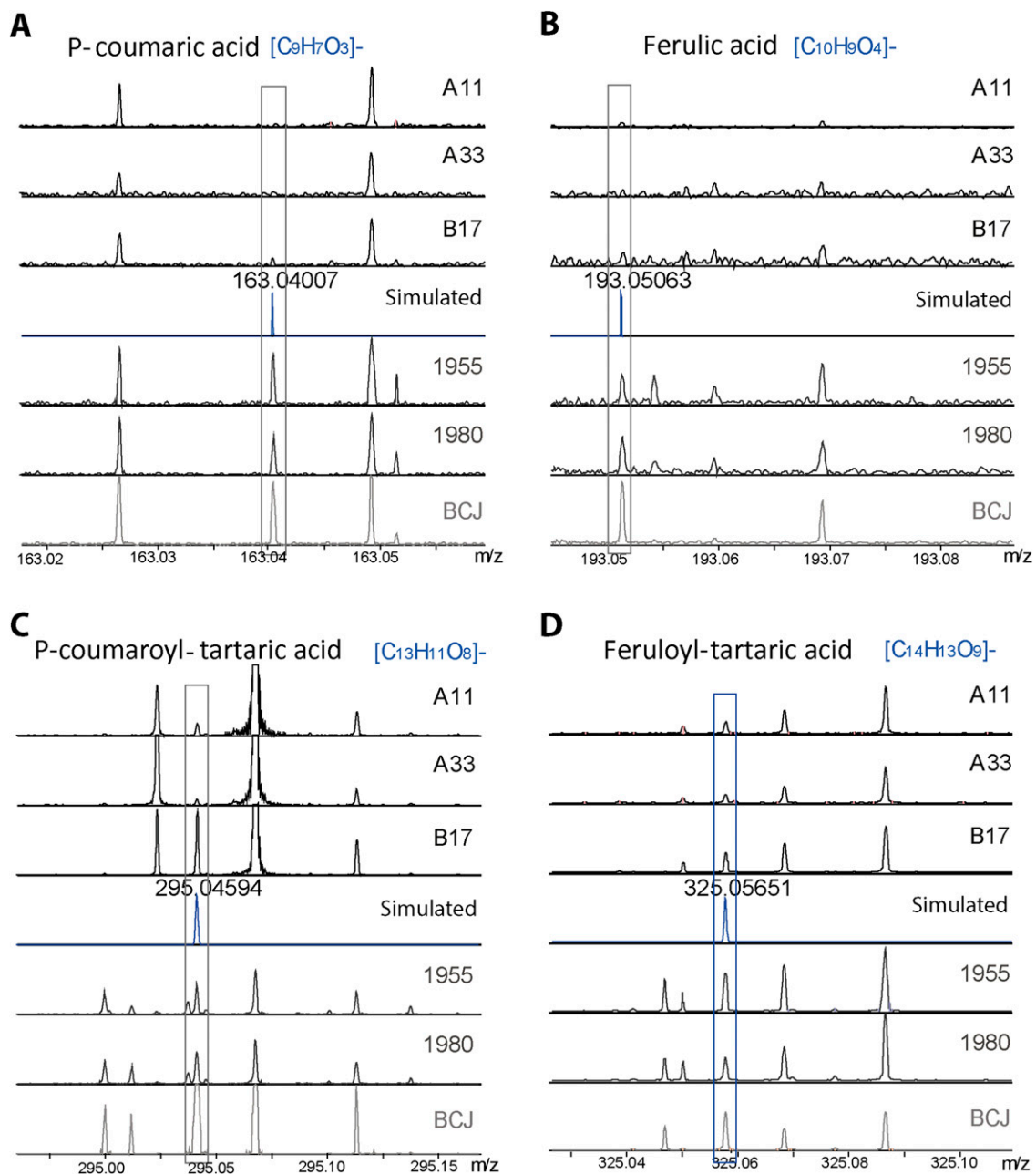


**Fig. S3.** Mass spectra of all champagnes. (A) In the  $m/z$  range from 120 to 550 normalized to tartaric acid signal intensity and showing the main signal differences. The most intense signals occur in Baltic Sea champagnes and can be assigned to malic acid, citric acid, tartaric acid, monosaccharides, and disaccharides; many of these signals correspond to chlorine adducts produced in the electrospray source due to the presence of chlorine ions. (B) In Baltic champagne, an increased intensity of chlorine adducts of Di-D-fructose dianhydrides  $[C_{12}H_{20}O_{10}Cl]^-$  and  $[C_{12}H_{22}O_{10}Cl]^-$  is noted. These are well-known markers of caramelization and thus give strong evidence of the addition of heated grape juice, in addition to the presence of HMF as shown in Fig. 2.



**Fig. S4.** Compositional network of 4,196 *m/z* features as obtained by FTICR/MS analysis of Baltic Sea and modern champagnes. (A) Color-differentiated representation of chemical classes highlighting, in B, one of the nitrogen-rich zones involving peptidic structures enriched in the modern champagne samples. (C) CHNO–sulfur compositional space. (D) CHNO–phosphorus compositional space. (E) Compound class enrichment analysis for modern and old champagnes showing the relative enriched structural characteristics in old (phosphorus) and modern champagnes (nitrogen compounds).





**Fig. S6.** FTICR/MS characterization of volatile phenol precursors. (A) Details of the mass spectra  $m/z$  163 showing the presence of the  $m/z$  163.04007 signal only in modern champagnes, assigned to  $[C_9H_7O_3]^-$ ; (B)  $m/z$  193 showing the presence of the  $m/z$  193.05063 signal only in modern champagnes, assigned to  $[C_{10}H_9O_4]^-$ ; (C)  $m/z$  295 showing the presence of the  $m/z$  295.04594 signal in all champagnes, assigned to  $[C_{13}H_{11}O_8]^-$ ; and (D)  $m/z$  325 showing the presence of the  $m/z$  325.05651 signal in all champagnes, assigned to  $[C_{14}H_{13}O_9]^-$ .



**Table S1. Current analyses of modern champagnes vs. champagnes from the Baltic Sea**

Measured parameter	Modern champagnes	Champagnes from the Baltic Sea		
		A11	B17	A33
Alcohol, % V/V	12.33	9.34	9.84	9.40
Density at 20 °C	0.990	1.032	1.022	1.032
pH	3.03	3.16	3.16	3.14
Total acidity, g H <sub>2</sub> SO <sub>4</sub> /L	4.7	4.2	4.1	4.2
Acetic acid, g/L	0.23	0.41	0.54	0.38
Malic acid, g/L	0.1	1.6	1.3	1.6
Tartaric acid, g/L	3.5	1.8	1.3	1.8
Sugars, g/L	1.6	144	118	144
Gluconic acid, mg/L	60	156	110	155

**Table S2. Elemental analyses of modern champagnes vs. champagnes from the Baltic Sea**

Element	Modern champagnes			Champagnes from the Baltic Sea		
	BCJ	1980	1950	A11	B17	A33*
Al	726	1,220	1,120	2,870	2,550	3,710
B	2,630	2,820	3,310	2,250	2,530	3,380
Ba	36	45	65	71	170	149
Br	nd	nd	nd	4,020	2,080	2,390
Ca	83,000	58,200	57,200	58,400	63,700	85,700
Cl	6,030	7,200	12,500	966,000	925,000	1,540,000
Cu	27	40	78	100	303	1,430
Fe	1,100	3,630	4,630	118,000	84,200	13,300
K	297,000	390,000	505,000	359,000	280,000	489,000
Mg	64,600	70,300	58,400	79,900	84,300	129,000
Mn	743	1,070	847	922	1,100	1,370
Na	8,020	18,700	9,480	510,000	443,000	1,050,000
NO <sub>3</sub> <sup>-</sup>	1,550	2,160	10,400	7,200	10,600	513,000
P	118,000	110,000	73,000	92,200	82,600	113,000
PO <sub>4</sub> <sup>3-</sup>	263,000	237,000	168,000	150,000	142,000	210,000
S	104,000	120,000	168,000	88,800	77,600	165,000
SO <sub>4</sub> <sup>2-</sup>	130,000	166,000	315,000	131,000	228,000	189,000
Sr	217	292	261	537	545	845
Ti	25	101	24	16	271	256
V	19	54	19	22	48	94
Zn	574	989	721	422	1,020	944
Elements closed to their limits of detection						
As	82	83	82	207	416	415
Cd	5	8	8	10	26	39
Co	26	13	13	51	127	199
Cr	17	20	17	20	50	87
Hg	8	9	8	22	45	44
Li	6	8	4	11	29	41
Mo	24	16	12	47	116	225
Ni	16	21	17	21	53	81
Pb	27	67	575	554	357	508
Sn	38	38	38	96	193	193

Concentrations are given in micrograms per liter; nd, not detected. The striking differences in Cl and Na concentrations between Baltic champagnes and modern ones straightforwardly suggest some seawater contamination of the former ones. However, several arguments can be made against such a hypothesis, even for the A33 sample, which was considered to be contaminated given the salty flavor detected during tasting. Assuming that A33 was contaminated, the relative proportion of sea-derived compounds in this wine is an indication of expected concentrations resulting from contamination from the Baltic Sea, the composition of which is known to be consistent throughout open seas and oceans. If all of the samples were contaminated, the relative proportion of Cl and Br that was found in A33 (Cl/Br = 645) should be found in champagnes A11 and B17. However, the Cl/Br ratios for these two champagnes are 444 and 240, respectively, which are significantly different from that of A33. The same conclusion can be drawn from Cl/SO<sub>4</sub><sup>2-</sup> ratios (around 8, 4, and 7 for A33, B17, and A11, respectively). Furthermore, a possible seawater contamination raises the question of diffusion, either through the cork or along defects at the cork/bottle interface, on the basis of the osmosis from high to low concentrated media. In the particular case of K, it is expected that the wine concentration is initially higher than the Baltic Sea one. Consequently, the K concentration of A33 should reflect an equilibrated concentration between the wine and the sea. In that case, the lower K concentrations for champagnes A11 and B17 are not consistent with seawater contamination, unless these three samples were stopped with corks having different permeability and selectivity. No literature data can be found to evaluate this hypothesis. Finally, the extremely high NO<sub>3</sub><sup>-</sup> concentration of A33 is certainly the most difficult to explain without considering, indeed, some contamination of this particular bottle.

\*Seawater-contaminated sample.

**Table S3. Relative determinations of metabolites in champagnes by <sup>1</sup>H NMR spectroscopy**

Metabolite*	Champagnes from the Baltic Sea			Modern champagnes			Average		t test
	A11	B17	A33	BCJ	1980	1955	Baltic champagnes	Modern champagnes	
Glucose	279.5	195.2	430.3	16.1	12.2	6.7	301.6	11.7	0.015
Fructose <sup>†</sup>	120.3	120.1	116.9	18.6	35.6	19.1	119.1	24.4	<0.001
Tartaric acid	7.6	5.5	9.6	13.3	12.9	15.3	7.6	13.8	0.011
Malic acid	10.8	7.8	13.8	1.3	2.2	2.3	10.8	2.0	0.007
Lactic acid	13.3	17.0	17.9	32.8	30.1	30.7	16.1	31.2	0.032
Acetic acid	3.30	3.96	3.66	1.96	4.26	4.44	3.6	3.6	0.919
Formic acid	0.6	0.3	0.7	0.1	0.2	0.1	0.5	0.1	0.022
Succinic acid	0.1	0.1	0.1	0.1	0.0	0.1	0.1	<0.1	0.823
Gallic acid	0.0	0.0	0.0	0.1	0.2	0.1	0.0	0.1	0.032
Ethanol	416.4	792.1	768.5	923.1	1137	804.7	659.0	955.1	0.13
2,3-Butanediol	3.6	2.7	3.0	4.0	5.7	6.2	3.1	5.3	0.038
Hydroxymethylfurfural	0.4	0.4	0.5	0.0	0.0	0.0	0.4	0.0	<0.001
Ratio malic acid/lactic acid	0.81	0.46	0.77	0.04	0.07	0.07	0.68	0.06	0.005
Glucose + fructose	399.8	315.3	547.2	34.7	47.8	25.8	420.8	36.1	0.005

\*Values are based on area integration of <sup>1</sup>H NMR peaks, normalized to the added standard TSP and stoichiometrically corrected for the number of protons.

<sup>†</sup>The fructose signal at δ4.12 showed overlap with lactate (HO-CH-CH<sub>3</sub>), and values might therefore be biased.

**Table S4. Identified volatile compounds of Baltic champagnes showing concentrations higher than their detection threshold**

Compounds	Mean, µg/L rel Kow	Error, %	Threshold, µg/L	Descriptors	Chemical family
Propylphenol	177,692	5	0.29*	animal note cowshed	volatile phenols
Isoamyl alcohol	176,192	3	30	alcoholic, malty, fusel	higher alcohols
2-Phenylethanol	21,309	3	60	rose, floral, honey	higher alcohols
Ethyl pentanoate	390	3	1.5	fruit, apple, floral	ethyl esters
5-Methylfurfural	1,771	4	16	caramel, spicy	Maillard product
Diethyl succinate	97,795	4	1,200	fruit, wine, wet	ester
Eugenol	277.25	4	5	clove, spicy, sweet	volatile phenols
Ethyl hexanoate	152.63	5	3	fruit, green apple, sweet	ethyl esters
Ethyl lactate	491,607	1	14,000	fruit, raspberry, perfume	ethyl esters
Butyric acid	5,244	8	173	rancid, cheesy, sweaty	medium-chain fatty acid
1-Hexanol	2,605	4	110	green, floral, resinous	alcohols
Ethyl dihydrocinnamate	32.85	2	1.60	floral, fruit, sweet	ester
1-Heptanol	58.80	5	3	green, fruit	alcohols
<i>trans</i> -Oak lactone	13,346	7	790	spicy, oak	lactone
Ethyl-2-hydroxy-2-methyl butanoate	11,326	8	1,000	floral, pineapple,	ethyl esters
Isoamyl lactate	1,701	6	200	unknown	esters
Isoamyl acetate	251	6	30	banana, fruit, sweet	esters
1-Octanol	782	2	110	green, floral, rose	alcohols
Hexanoic acid	2,219	44	420	cheesy, rancid, sweaty	medium-chain fatty acid
Ethyl octanoate	8.78	12	2	fruit, sweet, floral	ethyl esters
<i>cis</i> -Oak lactone	154.76	18	46	spicy, maple	lactone
Octanoic acid	1,587	12	500	fatty, rancid	medium chain fatty acid
Methyl eugenol	25.08	5	10	strawberry	volatile phenols
<i>cis</i> -3-Hexenol	134.59	10	70	green, grassy, fresh	alcohols
Eucalyptol	3.52	12	3.20	minty, sweet, spicy	terpenes
Isobutanol	15,566	48	1,000	wine, solvent	higher alcohol

All compounds were identified by Mass Spectrometry and Linear Retention Index (KOVATS) (see Table S5). rel Kow, related to the Kovats index.

\*Thresholds are calculated in wine, except for water.

**Table S5. KOVATS indexes and techniques used for aroma identification**

Compounds	Experimental KOVATS indexes	KOVATS indexes range In the literature	Identification*
Propylphenol	2,240	2,251	MS, LRI
Isoamyl alcohol	1,184	1,168–1,342	MS, LRI, std
2-Phenylethanol	2,176	1,915–2,262	MS, LRI, std
Ethylpentanoate	1,136	1,116–1,170	MS, LRI, std
5-Methylfurfural	1,375	1,366–1,411	MS, LRI
Diethylsuccinate	1,669	1,666–1,750	MS, LRI, std
Eugenol	2,151	1,994–2,237	MS, LRI, std
Ethyl lactate	1,317	1,311–1,374	MS, LRI, std
Butyric acid	1,816	1,853	MS, LRI, std
1-Hexanol	1,328	1,319–1,398	MS, LRI, std
Ethylidihydrocinnamate	1,880	1,353–1,908	MS, LRI, std
1-Heptanol	1,447	1,425–1,481	MS, LRI, std
<i>trans</i> -Oak lactone	1,887	1,862–1,987	MS, LRI
Ethyl-2-hydroxy-2-methyl butanoate	1,419	1,397–1,431	MS, LRI
Isoamyl lactate	1,563	1,555–1,619	MS, LRI, std
Isoamyl acetate	1,128	1,160–1,363	MS, LRI, std
1-Octanol	1,654	1,619–1,830	MS, LRI, std
Hexanoic acid	1,859	1,820–1,897	MS, LRI, std
Ethyl octanoate	1,432	1,392–1,497	MS, LRI, std
<i>cis</i> -Oak lactone	1,959	1,886–2,135	MS, LRI
Octanoic acid	2,060	2,001–2,120	MS, LRI, std
Methyl eugenol	2,002	1,923–2,099	MS, LRI
<i>cis</i> -3-Hexenol	1,365	1,320–1,415	MS, LRI
Eucalyptol	1,209	1,115–1,232	MS, LRI, std
Isobutanol	1,103	1,058–1,252	MS, LRI, std

\*MS, LRI, identified on the basis of both mass spectral data and Linear Retention Index data (in-house database and refs. 7–9) and/or by comparison with an authentic standard (std).