



Identificació de marcadors genètics i fenotípics per a les tipologies de Tomata Valenciana utilitzant l' Atlas de Tomata Tradicional Europea

Identification of genetic and phenotypic markers for Tomata Valenciana Typologies using the Tradional European Tomato Atlas

Clara Pons^a, Savador Soler^b, Antonio J. Monforte^a i Antonio Granell^a

^aInstituto de Biología Molecular y Celular de Plantas, (CSIC-UPV), cpons@upvnet.upv.es, ^bInstituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana.

How to cite: Pons, C.; Monforte, A.J.; Granell, A.; Pérez-Martínez, J.A. 2024. Identificació de marcadors genètics i fenotípics per a les tipologies de Tomata Valenciana utilitzant l' Atlas de Tomata Tradicional Europea. En libro de actas: *II Congrés de la Tomata Valenciana. L'Autèntica*. València, 30 de maig de 2024. <https://doi.org/10.4995/TOMAVAL2024.2024.18684>

Abstract

Southern Europe is a secondary center of tomato diversification, especially in the Mediterranean basin, with a wide variety of types adapted to the environment. Recent studies have revealed extensive genetic and phenotypic variability, mainly related to fruit shape and color. In the Valencia region, the 'Valenciana' tomato is one of the most well-known and valued varieties, with three main typologies based on morphological characteristics. As visual classification is prone to errors, it is necessary to develop robust and reliable methods for classifying 'Valenciana' tomato accessions such as those based on genetic and phenotypic traits. In this work, by MFA integrating phenotypic and genotypic data followed by HCPC analysis, we have produced molecular and phenotypic signatures for each type of typology, allowing for a reliable and objective classification of the different typologies.

Keywords: *Tomato 'Valenciano', phenotype, genotype, data mining, data Integration, classification*

Resum

El sud d'Europa és un centre secundari de diversificació de la tomata, especialment a la conca Mediterrània, amb una gran varietat de tipus adaptats al medi. Estudis recents han revelat una àmplia variabilitat genètica i fenotípica, principalment relacionada amb la forma i el color del fruit. A la regió de València, la tomata 'Valenciana' és una de les

varietats més conegudes i apreciades, amb tres tipologies principals basades en característiques morfològiques. Donat el risc d'errors associat a la classificació visual, és necessari buscar mètodes fiables per a la classificació de les accessions de tomata 'Valenciana' basades en característiques genètiques i fenotípiques. En aquest treball, mitjançant la integració de dades fenotípiques i genotípiques amb un anàlisi MFA seguit de HCPC, hem identificat signatures moleculars i fenotípiques per a cada tipus de tipologia, permetent una classificació fiable i objectiva de les diferents tipologies

Paraules clau: Tomata 'Valenciana', fenotip, genotip, mineria de dades, integració de dades, classificació

1. Introducció

Des de la seua arribada a Europa (a principis del segle XVI), la tomata va ser ràpidament adoptada a les dietes espanyoles i italianes (Gentilcore, 2009; López-Terrada, 2017) i, posteriorment, des del segle XVIII, a les cuines de la resta de països europeus, i posteriorment a la resta del món [3]. El sud d'Europa és un centre secundari de diversificació de la tomata (Blanca 2022), sent la conca Mediterrània una de les regions amb major diversitat en varietats locals (Mazzucato et al., 2008). En aquestes regions s'han desenvolupat un gran nombre de varietats locals de consum en fresc, llarga vida i processat adaptades al medi on es cultiven, amb fruits de diferents formes i colors. Estudis recents en una col·lecció de més de 1500 tomàquets tradicionals europeus (Blanca et al., 2022; Pons et al., 2023, 2022) han revelat una àmplia variabilitat genètica i fenotípica en la tomata tradicional Europa. Aquesta diversificació sembla que ha estat especialment dirigida per trets relacionats amb la forma i el tamany del fruit, principalment associats a l'ús culinari diferent en cada país i regió (Pons et al., 2022). A banda, aquest estudi també ha posat de manifest que el color del fruit sembla estar més relacionat amb les preferències "estètiques" en cada país, que amb el seu ús (Pons et al., 2022).

A la regió de València, les principals varietats tradicionals de tomata són les anomenades 'de borseta', 'del pebre', 'de penjar', 'de pruna', 'Mutxamel' i 'Valenciana' (Cebolla-Cornejo, Roselló, & Nuez, 2013; Figàs, Prohens, Casanova, Fernández-de-Córdova, & Soler, 2018). Entre elles, la tomata 'Valenciana' és una de les varietats més conegudes i apreciades per les seues propietats organolèptiques (Soler, 2024). Als últims anys, el interès per les tomates tradicionals augmenta cada vegada més, promocionat per un consum de producte local, i la creença de que les varietats locals solen tenir propietats nutritives i sensorials superiors ("una creença de "sabor del passat") (Casañas, Simó, Casals, & Prohens, 2017). El cas de la tomata 'Valenciàna' és un exemple de l'interès creixent per les varietats tradicionals de tomata. En els últims deu anys, ha mostrat una projecció econòmica molt notable, amb una producció comercialitzada anual superior a 1.250.000 kg i un valor d'aquesta producció de més de dos milions d'euros (Soler, 2024). Al

respecte, és vital la realització d'estudis sobre la diversitat que presenten les varietats locals d'aquest tipus de tomata i com conservar-les mitjançant el seu cultiu (Figàs et al., 2018).

Els fruits de tomata 'Valenciana' són, en general, fruits de gran mida, amb lòculs petits distribuïts de forma irregular al voltant del cor de secció circular. A partir de característiques com la intensitat de l'espalla del fruit, el color del fruit immadur i l'apuntament del fruit, s'han establert tres tipologies de Tomata Valenciana: *Masplet*, *Blanca* i *Intermedia* (Soler, 2024). El tipus *Masplet* es caracteritza per presentar un gran apuntament, menor amplada de fruit, uns muscles verd intens i un color de fruit immadur verd fosc. Pel que fa al tipus *Blanca*, té una forma de cor més aplanada, una part posterior sovint menys pronunciada i un color verd més atenuat o blanquinos en estadiu immadur, el que dóna nom al tipus. El fruit madur té un color vermell intens. Generalment, tant el tipus *Masplet* com el *Blanca* tenen lòculs petits en gran nombre, que en el tipus *Masplet* es troben més agrupats, de forma més regular i en menor nombre que en la tipologia *Blanca*. Finalment, la tipologia *Intermedia* mostra característiques típiques del tipus *Masplet*, i altres són típiques del grup *Blanca*. Aquest tipus presenta una major amplada de fruit, menor coloració verda als muscles, menor apuntament o nul, més pes i fruit immadur de color més clar. A banda, podríem incloure un quart tipus, la tomata 'Valenciana' *Rosa*, que presenta característiques semblants a la blanca, però de coloració rosa.

No obstant això, aquesta caracterització generalment es fa visualment, el que dificulta identificar de manera específica i objectiva les diferents tipologies. El coneixement de les característiques fenotípiques, genètiques i bioquímiques distintives de cada tipologia, així com la seua diversitat i estructura genètica, són claus tant per a la seua valorització i conservació com per desenvolupar estratègies òptimes per a la seua millora.

Recentment, hem desenvolupat un Atlas de diversitat fenotípica, genotípica i geogràfica present en la tomata tradicional Europea (Pons et al., 2022), utilitzant una col·lecció de 1489 accessions de tomata (col·lecció TRADITOM). Aquesta col·lecció l'hem usada per desvetllar les bases moleculars de la diversitat fenotípica de la tomata que es cultiva, o s'ha cultivat, al sud d'Europa als últims 70 anys. Aquest treball, el més gran realitzat fins el moment en tomata, ens ha permès identificar els principals tipus agromorfològics de tomata cultivats al sud d'Europa. A banda, mitjançant un meta-anàlisi d'associació de genoma ampli (GWAS), vam identificar i validar creuadament 1486 associacions per a 70 trets en 211 loci, diversos dels quals amb potencials efectes pleiotròpics. Finalment, en aquest estudi, vam identificar signatures moleculars i combinacions de loci que van revelar diferents històries genètiques i els fonaments de la variació fenotípica de diferents tipus de tomata al sud d'Europa.

Amb l'objectiu de poder identificar característiques fenotípiques i genètiques distintives per a definir i classificar correctament les accessions de tomata 'Valenciana' en aquest treball hem fet mineria de dades en l'Atlas de Tomata Tradicional. En aquest treball, proporcionem una sèrie de marcadors genètics i fenotípics necessaris per a poder diferenciar objectivament entre les

diferents tipologies de tomata 'Valenciana', que seran de gran utilitat per a la seua classificació, valorització i conservació, així com per desenvolupar estratègies òptimes per a la seua millora.

2. Materials i mètodes

2.1. Dades

Fer a aquest estudi es van utilitzar les dades fenotípiques i genètiques generades a l'Atlas de Tomata Tradicional (Pons et al., 2022). A les dades es pot accedir directament des de la pagina web <https://www.gateway.harnesstom.eu/>. Per a les dades fenotípiques es va utilitzar les dades del estudi PERELLO_summer15 (<https://www.gateway.harnesstom.eu/study/>) generades als camps experimentals del Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana. La matriu genotípica correspon a Variant_set_TRADITOM.vcf (<https://www.gateway.harnesstom.eu/variant/set/metadata/>). Els fitxers es van filtrar per obtenir les dades fenotípiques i genètiques de 53 tomates 'Valencianes' (Taula 1) amb les que es va treballar posteriorment.

Taula 1. Procedència de las entrads de Tomata 'Valenciana' en la col·lecció TRADITOM i tipologia assignada visualment a les dades de passaport

Tipologia	Accesió	Lloc recolecció	Província	Regió
BLANCA	TRVA2340	Cabanes	Castelló	Comunitat Valenciana
BLANCA	TRVA3220	Toga	Castelló	Comunitat Valenciana
BLANCA	TRVA2350	Vilanova d'Alcolea	Castelló	Comunitat Valenciana
BLANCA	TRVA2940	Agullent	València	Comunitat Valenciana
BLANCA	TRVA2950	Albalat de la Ribera	València	Comunitat Valenciana
BLANCA	TRVA1880	Albuixech	València	Comunitat Valenciana
BLANCA	TRVA2630	Alginet	València	Comunitat Valenciana
BLANCA	TRVA2570	Almussafes	València	Comunitat Valenciana
BLANCA	TRVA1250	Chelva	València	Comunitat Valenciana
BLANCA	TRVA1940	Cullera	València	Comunitat Valenciana
BLANCA	TRVA0020	El Perelló	València	Comunitat Valenciana
BLANCA	TRVA1920	El Perelló	València	Comunitat Valenciana
BLANCA	TRVA2580	El Perelló	València	Comunitat Valenciana
BLANCA	TRVA2610	El Perelló	València	Comunitat Valenciana
BLANCA	TRVA3130	El Puig	València	Comunitat Valenciana
BLANCA	TRVA1930	Foios	València	Comunitat Valenciana
BLANCA	TRVA3140	La Pobla de Vallbona	València	Comunitat Valenciana
BLANCA	TRVA2070	L'Alcúdia	València	Comunitat Valenciana
BLANCA	TRVA3050	L'Elinana	València	Comunitat Valenciana
BLANCA	TRVA2330	Macastre	València	Comunitat Valenciana
BLANCA	TRVA1960	Moncada	València	Comunitat Valenciana
BLANCA	TRVA1970	Moncada	València	Comunitat Valenciana
BLANCA	TRVA2320	Museros	València	Comunitat Valenciana
BLANCA	TRVA3080	Naquera	València	Comunitat Valenciana
BLANCA	TRVA2050	Paterna	València	Comunitat Valenciana
BLANCA	TRVA3100	Picanya	València	Comunitat Valenciana
BLANCA	TRVA2590	Picassent	València	Comunitat Valenciana
BLANCA	TRVA3110	Pinedo	València	Comunitat Valenciana
BLANCA	TRVA2620	Quatretonda	València	Comunitat Valenciana
BLANCA	TRVA3240	Torres Torres	València	Comunitat Valenciana
BLANCA	TRVA1800	Turís	València	Comunitat Valenciana
BLANCA	TRVA3250	Turís	València	Comunitat Valenciana
BLANCA	TRVA3260	Utiel	València	Comunitat Valenciana
BLANCA	TRVA1860	València	València	Comunitat Valenciana
BLANCA	TRVA2310	Villargordo del Cabriel	València	Comunitat Valenciana
BLANCA	TRVA0990	Jiménez de Jamuz	León	Castilla y León
INTERMITJA	TRVA1850	Moncada	València	Comunitat Valenciana
INTERMITJA	TRVA2120	Paterna	València	Comunitat Valenciana
INTERMITJA	TRVA3190	Silla	València	Comunitat Valenciana
INTERMITJA	TRVA1620	València	València	Comunitat Valenciana
INTERMITJA	TRVA2360	València	València	Comunitat Valenciana
INTERMITJA	TRVA2180	Cobatillas	Murcia	Murcia
MASCLET	TRVA3200	Sant Joan d'Alacant	Alacant	Comunitat Valenciana
MASCLET	TRVA3120	Puçol	València	Comunitat Valenciana
MASCLET	TRVA3180	Sagunt	València	Comunitat Valenciana
MASCLET	TRVA3210	Sueca	València	Comunitat Valenciana
MASCLET	TRVA3230	Torrent	València	Comunitat Valenciana
MASCLET	TRVA1950	Vinalesa	València	Comunitat Valenciana
VAL-ROSA	TRVA2920	Agullent	València	Comunitat Valenciana
VAL-ROSA	TRVA2970	Bocairent	València	Comunitat Valenciana
VAL-ROSA	TRVA3010	Castellonet de la Conquesta	València	Comunitat Valenciana
VAL-ROSA	TRVA3170	Set Aigües	València	Comunitat Valenciana

2.2. Anàlisi multifactorial i l'agrupació jeràrquica sobre components principals fenotípic

El anàlisi multifactorial (MFA) i l'agrupació jeràrquica sobre components principals (HCPC, de l'angles hierarchical clúster of principal componets) es van realitzar utilitzant el paquet Factominer de R (Husson, Josse, & Pages, 2010) els paquets de R FactoMiner i Factoextra (Kassambara, 2017; Lê, Josse, & Husson, 2008). Per al MFA les variables contínues es van escalar i es van seleccionar les configuracions estàndard del programa. Totes les variables es van establir com a actives. Les variables es van representar en el pla descrit per les dimensions principals del MFA. Les correlacions quadrades entre les variables (o grup de variables) i les dimensions es van utilitzar com a coordenades. Per definir els grups de variables o les variables més importants per explicar la variabilitat en el conjunt de dades, es va calcular la contribució total d'un grup de variables o una variable en explicar la variació retinguda per les n -dimensions ($C_{j,n}$) (Kassambara, 2017). El punt de tall de la contribució per a una variable j es va calcular suposant una contribució uniforme de la variable, de manera que el valor de la contribució esperada per a la variable j a la dimensió i és $1/Y$, on Y és el nombre total de variables.

Taula 2. Grups de variables i variables fenotípiques utilitzades en aquest estudi

Grup	Variables	Tipus de variable
<i>aparença externa</i>	Fasciació del fruit, nervadura a la part de calix	qualitativa
<i>aparença interna</i>	Oquetat del fruit	qualitativa
<i>aparença dels muscles</i>	Forma i color dels muscles	qualitativa
<i>tipus</i>	Tipologia tomata 'Valenciana'	qualitativa
<i>mida</i>	Area, amplària màxima, amplària a mitat llargaria, amplària a la posició més ampla, pes del fruit, llargaria a mitad d'amplària, màxima llargaria	quantitativa
<i>forma distal</i>	Angle distal macro, angle distal micro, afonament del la part distal, index de la area distal, excentritat distal	quantitativa
<i>forma proximal</i>	Angle proximal macro, angle proximal micro, tancament del la part proximal, index de la area proximal, excentritat proximal, llargaria dels muscles	quantitativa
<i>lòculs</i>	Nombre de lòculs, lobulació	quantitativa
<i>Index de forma</i>	Index de forma externa I, Index de forma externa II, Index de forma interna	quantitativa
<i>qualitat</i>	Fermesa del fruit, sòlids solubles ($^{\circ}$ Brix), color extern de fruit a l'espai CIELAB a, b L	quantitativa

L'agrupació jeràrquica sobre components principals (HCPC, de l'angles hierarchical clúster of principal componets) es va realitzar utilitzant el paquet Factominer de R (Husson et al., 2010) sobre els resultats de les primeres 5 dimensions del MFA, utilitzant la distància euclidiana com a mètrica de distància i el criteri de Ward com a algorisme aglomeratiu. El nombre òptim de

clústers es va seleccionar satisfent tres criteris: (i) el gran nombre de clústers que (ii) maximitzen la pèrdua relativa d'inèrcia (Husson et al., 2010) i (iii) maximitzen el nombre d'accessions per tipologia.

L'enriquiment de característiques que defineixen cada clúster es va obtenir calculant el valor de *v.test*, que compara la proporció d'accessions en el clúster que comparteixen una característica, en comparació amb la proporció global. Les característiques amb un valor de $p \leq 0.05$ es van considerar significatives. El signe del *v.test* indica si la proporció en la que es troba la variable en el clúster està enriquida (en el cas de les variables qualitatives) o la mitjana de la variable en el clúster és més gran que la mitjana en el conjunt de dades complet (en el cas de les variables quantitatives). La selecció de SNPs en base al MFA-HCPC es va fer tenint en compte aquells SNPs que estaven presents o absents a totes accessions dins d'un mateix clúster.

2.3. Anàlisi molecular

Per realitzar les anàlisis moleculars es va utilitzar una matriu amb les 53 accessions de tomata 'Valenciana' i 2204 SNPs (del anglès, single nucleotide polymorphism) amb un MAF > 0,01 i una presència de SNPs >70% en les accessions. El conjunt de tomata 'Valenciana' va ser analitzat amb els marcadors SNP seleccionats utilitzant Multidimensional Scaling Analysis (MDS). Primerament, es va calcular una matriu de distància Identity-By-State (IBS) entre parells de SNPs, i després es realitza l'anàlisi MDS. El filtrat i MDS es van fer utilitzant programa TASSEL v5.0. La visualització del MDS es va fer utilitzant el paquet de R plotly.

3. Resultats i Discussió

Aquest treball té com a objectiu identificar marcadors genètics i fenotípics per a una classificació precisa de les accessions de tomata 'Valenciana'. Per avaluar els trets que expliquen les diferències entre les diferents tipologies de tomata 'Valenciana', primer es va realitzar un anàlisi MFA.

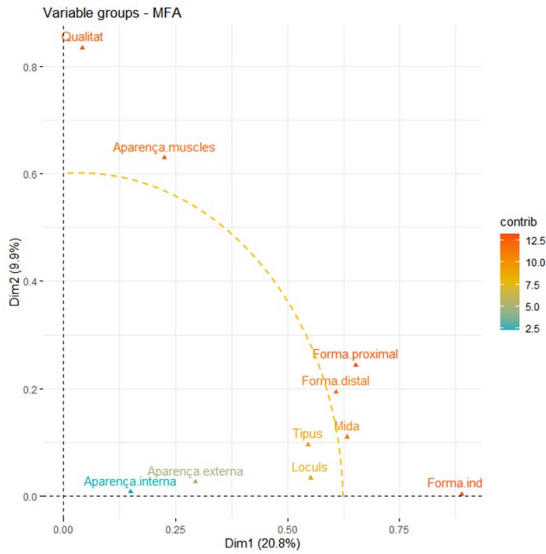


Figura 1. Mapa de factors variables que il·lustra la contribució i la relació de cada grup de variables a les dues primeres dimensions del MFA.

La línia taronja discontinua indica la contribució uniforme mitjana esperada (punt de tall)

Per l'anàlisi MFA les dades fenotípiques es van organitzar en deu grups de variables relacionades amb característiques de fruit, distribuïdes en quatre grups de variables qualitatives adquirides visualment (*aparença externa*, *aparença interna*, *aparença dels múscles* i *tipus*) i sis grups de variables quantitatives contínues (*mida*, *forma distal*, *forma proximal*, *nombre de loculs*, *Index de forma* i *qualitat*, adquirides amb aparell específics. Les característiques de cada grup de variables es detallen la Taula 2. Vam analitzar les dades fenotípiques juntament amb la tipologia: *Masclet*, *Blanca*, *Intermèdia* i *Rosa*.

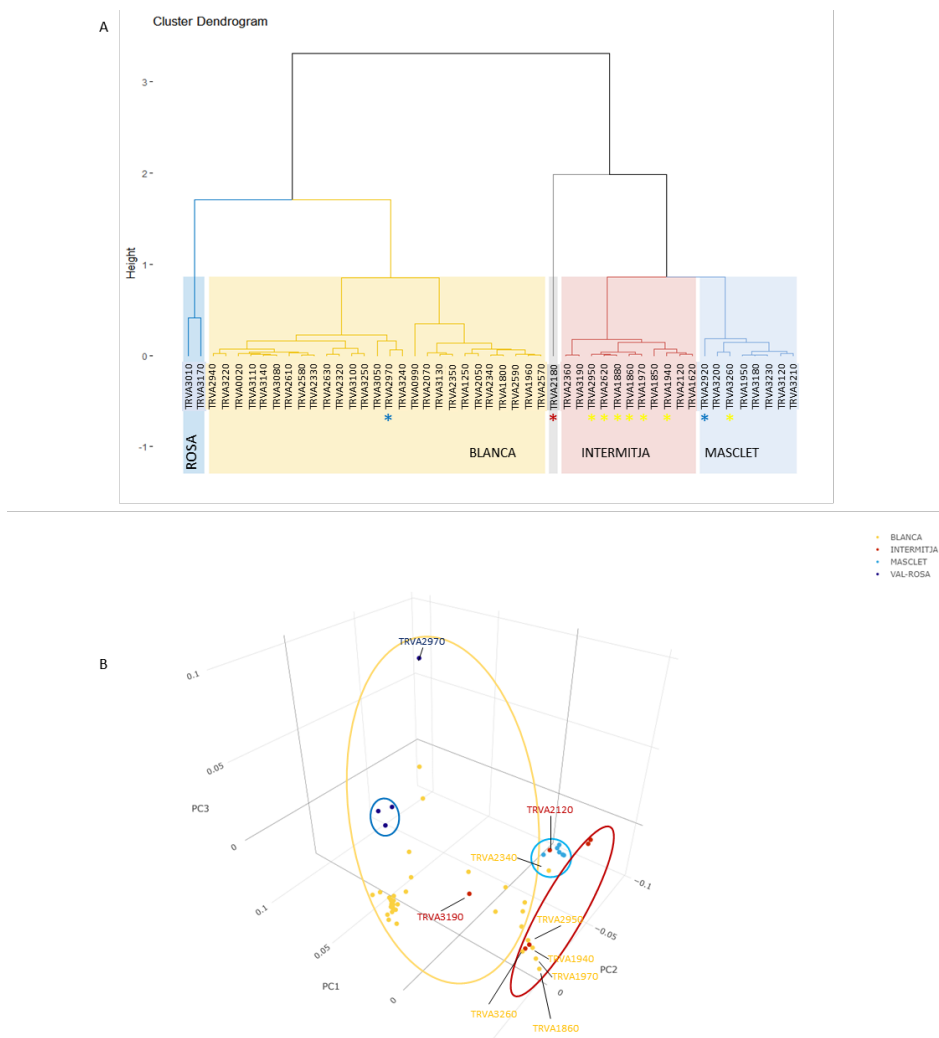


Figura 2. Classificació inicial de les tomates 'Valencianes' en base al fenotip i al genotip.

A) Dendrograma mostrant els cinc grups resultants del HCPC sobre les cinc primeres dimensions del MFA. Les imatges de tomata corresponen a la tomata més representativa de cada grup. Cada grup està representat per un color i la tipologia anotada en base a la tipologia predominant del grup. Les accessions marcades amb un asterisc indiquen accessions classificades visualment en el grup erroni, el color indica la tipologia assignada visualment. B) Anàlisi MDS de la col·lecció de tomata tradicional 'Valenciana' utilitzant els 2204 SNPs amb un MAF > 0,01. Les elipses indiquen les tipologies. El nom de l'accessió s'indica en el cas d'accessions classificades visualment en el grup erroni. Cada grup està representat per un color i la tipologia anotada en base a la tipologia predominant del grup.

Les contribucions de cadascun del grup de variables es mostren a la Figura 1. Els grups de trets que més contribueixen a les primeres dues primeres dimensions del MFA, i que per tant són les que globalment millor diferencien les tomates 'Valencianes' són (per ordre de contribució): forma proximal, index de forma, qualitat, aparença dels múscles, forma distal i mida. Com es pot veure la majoria de les variables que més contribueixen a diferenciar les accessions són aquelles que no estan obteses visualment.

Per poder classificar les 53 tomates valencianes de manera objectiva en base al fenotip es va realitzar un anàlisi HCPC utilitzant les cinc primeres dimensions del MFA. El anàlisi MFA amb HCPC va resultar en la classificació de les tomates 'Valencianes' en 5 grups (Figura 2A). Els grups es van anomenar en funció de quina era la tipologia predominant. El primer grup, contenia el 100% de les accessions etiquetades com a Masplet al passaport. El segon, estava compost per el 78% accessions etiquetades com Blanca. El tercer, conté el 83.3% de les accessions anotades com Intermèdia. El quart, conté el 50% de les accessions classificades com Rosa. Finalment, un grup format per una sola accessió classificada com Intermèdia. Les classificacions basades en el anàlisi MFA-HCPC van indicar que un 21% de les accessions etiquetades en el passaport com Blanca s'agrupaven junt a les intermèdies i una (TRVA3260) amb tipus Masplet (Figura 2A). A més, dues de les accessions etiquetades com Rosa, TRVA2970 i TRVA2920 s'agrupaven junt a les tipus Blanca i Masplet, respectivament.

De manera similar, el anàlisi MDS de les dades genotípiques amb un $MAF > 0.01$ (Figura 2B) va mostrar que algunes de les accessions no s'agrupaven amb la tipologia en la que s'havien etiquetat al passaport. La comparació en entre les accessions que agrupaven en el MFA-HCPC i MDS en grups diferents als que estaven etiquetats al passaport va indicar cert grau de solapament, així que les accessions es van reclassificar d'acord amb el fenotip. Les accessions reclassificades van ser sotmeses a un nou anàlisi MFA seguida de HCPC, on es van incloure les dades genotípiques amb un $MAF > 0.03$ com a variable addicional. El resultat va revelar l'existència de 6 grups, amb només algunes accessions que no concordaven amb les classificacions anteriors de MFA. Aquestes accessions van ser reassignades a la tipologia corresponent basant-se en aquesta nova classificació. Aquest procés de MFA-HCPC i reassignació es va repetir fins que totes les accessions van ser assignades a una de les quatre grups basant-se en el fenotip i el genotip, coincidint amb les etiquetes de les tipologies assignades (Figura 3).

Les característiques fenotípiques i genètiques que més diferencien les quatre tipologies es presenten a les Taules 3 i 4. L'anàlisi fenotípic dels clusters (Taula 3) indica que les tipologies ROSA i BLANCA tenen significativament major pes i fruits més amples, sent la ROSA, la que te els fruits més grans i més amples. Mentre que la tipologia INTERMITJA és la que te els fruits més petits i més estrets. A banda, les tipus MASPLET i INTERMITJA, tenen una llargaria significativament major que les altre dues tipologies. Les de tipologia MASPLET són les que

presenten major índexs de forma i major llargaria a mitad d'amplaria. Les *BLANQUES* i les *ROSA*, són però més aplanades i tenen índex de forma menors, especialment les *ROSA*.

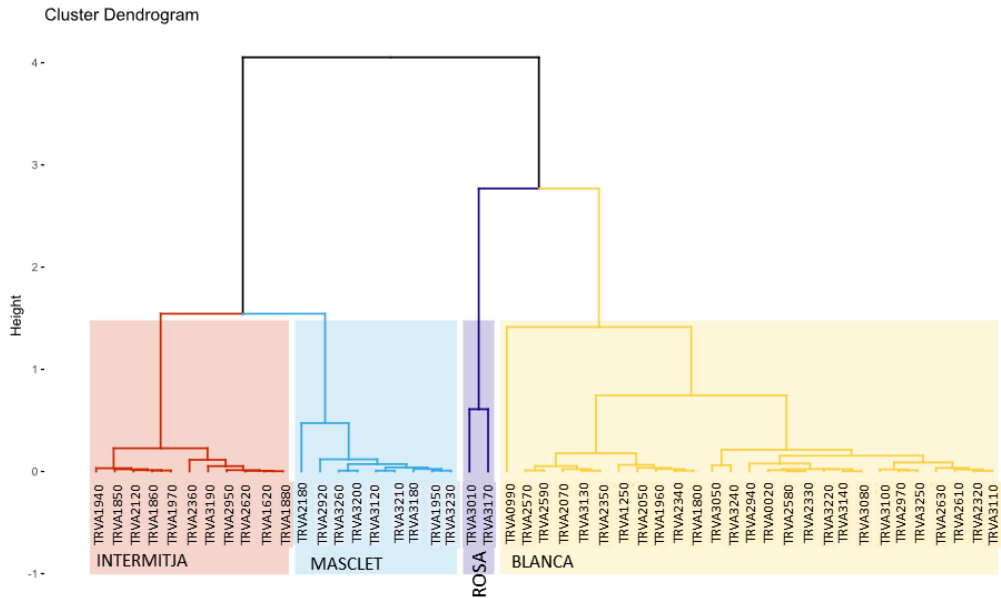


Figura 3. Classificació final de les tomates 'Valencianes' en base al fenotip i al genotip.

Dendrograma mostrant els quatre grups resultants del MFA- HCPC iteratiu. Les imatges de tomata corresponen a la tomata més representativa de cada grup. Cada grup, corresponent a una única tipologia esta representat per un color.

Respecte a les zones proximal i distal, les de la tipologia MASCLET tenen un angle distal macro i un tancament de la zona distal significativament menor que la resta, mentre que les de tipus BLANCA, INTERMITJA i BLANCA tenen aquests valors significativament majors. Indicant un major apuntament de la zona distal dels fruits tipus MASCLET, i menor als altres tipus. En el cas de la zona proximal, el anàlisi indica que les BLANQUES són les que tenen en promig major llargaria de muscles, però són més planes per la part proximal que les MASCLET i les INTERMITGES, que tenen un angle macro significativament menor. Aquest paràmetre, és màxim en el cas de les ROSA. A banda, els fruits ROSA es diferencien de la resta per tenir en promig un nombre significatiu major de lòculs i una major la lobulació. Finalment, els tipus MASCLET tenen major color roig (donat per baixos nivells del parametre de color a*), mentre que els tipus intermig tenen menor lluminositat del fruit (parametre L*) que la resta.

Taula 3. Diferenciació fenotípica de les quatre tipologies de tomata 'Valencians' en funció del MFA-HCPC. Les dades representen el valor promig dins de cada clúster i la desviació estandard (en el cas de variables quantitatives) o el percentatge d'accessions dins el clúster que compleixen aquest tret (en el cas de les qualitatives). Els nombres en roig representen aquelles variables que tenen una mitjana major que la mitjana global o estan sobre-representades. Els nombres en negre corresponen a valors en el clúster menors que la mitjana o infra-representats. ns: $p > 0.05$, no significant. Valors amb asteriscs *, **, ***, **** son estadísticament significatius a diferents valors de probabilitat de $p \leq 0.05$, ≤ 0.01 , ≤ 0.001 and ≤ 0.0001 respectivament.

Grup	Variable	MASCLET	INTERMITJA	BLANCA	VAL-ROSA
<i>Mida</i>	pes del fruit	215.55± 44.70*	199.76± 32.83**	306.16± 71.96*	554.31± 91.64****
	àrea		41.47± 3.959***	54.60± 9.1023**	69.42± 0.23**
	màxima llargaria		7.34± 0.39*		
	amplària màxima	7.47± 0.56*	7.31± 0.40**	8.89± 1.00***	10.74± 0.42**
	amplària a mitat llargaria	7.31± 0.58*	7.17± 0.39**	8.78± 1.00***	10.65± 0.40**
<i>Forma</i>	amplària a la posició més ampla	0.41± 0.02**	0.41± 0.01*	0.43± 0.02**	
	llargaria a mitat d'amplària	7.01± 0.38*			
	Angle distal macro	113.67± 12.88***		140.87± 13.64***	
<i>distal</i>	excentritat distal		0.88± 0.00***	0.89± 0.00*	0.89± 0.01**
	tancament del la part distal	0.47± 0.06***	0.51± 0.03*	0.58± 0.03****	
	índex de la area proximal	0.07± 0.04***			
<i>Forma</i>	Angle proximal macro	197.01± 33.19*	187.95± 24.33***	233.60± 19.66***	285.73± 20.87**
	angle proximal micro	248.04± 20.41*			
	tancament del la part proximal	0.71± 0.05**		0.77± 0.03**	
<i>Forma</i>	índex de la area proximal		0.10± 0.03*		
	llargària dels muscles		0.05± 0.01***	0.06± 0.01**	
	Índex de forma externa I	1.07± 0.06****	1.01± 0.04*	0.89± 0.06****	0.78± 0.06*
<i>Index</i>	Índex de forma externa II	0.97± 0.07***	0.93± 0.05**	0.77± 0.07****	0.63± 0.08*
	Índex de forma interna	0.98± 0.06***	0.94± 0.05**	0.79± 0.06****	0.65± 0.08**
<i>Loculs</i>	lobulació		1.18± 0.19*		2.42± 0.03***
	Nombre de loculs				17.55± 7.44***
<i>Qualitat</i>	color a	19.02± 2.27*			
	color L		33.17± 1.81*		
<i>aparença externa</i>	Fasciació del fruit baixa		9.09%*	59.26%****	
	Fasciació del fruit absent	100%**		29.63%***	
<i>aparença dels muscles</i>	Forma dels muscles lleugerament			14.81%*	
	Color dels múscles verd pàlid			37.04%*	
	Color dels múscles verd internig			62.96%*	
<i>aparença interna</i>	Oquetat del fruit baixa				50%*
	Oquetat del fruit intermitja			33.33%*	

Respecte al genotip (Taula 4) hi ha 335 SNPs que presenten diferències significatives (p -valor ≤ 0.05) en les freqüències al·lèliques a les diferents tipologies. No obstant, no hi ha cap SNP que distingisca els quatre tipus varietals, més aviat, la diferència es deu a la combinació d'ells. Només en el cas de 151 SNPs es detecten freqüències del 0% o 100% en alguna tipologia. Així, per exemple, hi ha SNPs en els quals el 100% de les accessions d'una tipologia només presenta

un al·lel, dels dos possibles que pren el SNP, en homozigosi, o casos en els quals el que caracteritza a la tipologia és el fet que un dels al·lells no apareix mai en homozigositat o heterozigositat en aquesta tipologia, però sí l'altre al·lel o l'heterozigot, o viceversa. La majoria dels SNPs amb el 100% de les accessions amb un al·lel dels possibles que pren el SNP en homozigosi es troben a la tipologia ROSA. No obstant, en aquest cas només hi ha dues accessions al clúster i, per tant, és més fàcil trobar SNPs que siguin iguals en totes les accessions del clúster.

D'aquests 151, hi ha 73 SNPs per als quals a l'Atlas de tomata tradicional (Pons et al., 2022) hi ha una associació amb algun caràcter de forma o color (Taula4). Així per exemple, el SNP SL2.50ch01p78823494 a la tipologia INTERMITJA no pren mai el genotip AA. La presència del genotip AA en aquest SNP està relacionada amb el color rosa dels fruits i una pell sense color en tomata tradicional (Pons et al., 2022). Aquest SNP té com a gen candidat *Sl_MYB12* (y), reponsable de la falta de color de la pell del fruit i del color rosa (Ballester et al., 2010). Un altre exemple, és el cas l'absència del genotip GG al SL2.50ch08p3352366, que caracteritza a les tipologies MASCLET i INTERMITJA. L'absència de l'al·lel G en homozigosi s'ha associat amb la forma oxheart (Pons et al., 2022) i aquest SNP colocalitza al mateix locus que *SISUN21*, un membre d'una família multigènica relacionada amb l'augment de la llargària del fruit (Huang, Van Houten, Gonzalez, Xiao, & van der Knaap, 2013). Un altre cas per exemple és el SNP SL2.50ch07p66971471 que a les tipus BLANCA mai presenta l'al·lel G en homozigosi, mentre que al tipus VAL-ROSA, en les dues accessions apareix el genotipe GG. La presència del al·lel G s'ha associat amb l'augment de l'amplària màxima del fruit, l'amplària a la mitdad de longitud i la fasciació del fruit (Pons et al., 2022). Aquest SNPs es troba prop del gen *flacca* que s'ha relacionat, entre altres, amb el pes del fruit, el grossor del pericarp (Nitsch et al., 2012; Sagi, Scazzocchio, & Fluhr, 2002)

Taula 4. SNPs que diferencien significativament (p -valor ≤ 0.05) les tipologies de tomata 'Valenciana' i, amb al·lells que es troben sempre presents o sempre absents en les diferents tipologies.

Els SNPs marcats amb un asterisc es troben associats a algun tret per GWAS a l'Altes de Tomata tradicional (Pons et al., 2022)

SNP	Allels	MASCLET	INTERMITJA	BLANCA	VAL-ROSA
SL2.50ch01p78823494*	G/A		GG/G:A, no AA		
SL2.50ch01p78965718*	C/T		CC/C:T, no TT		
SL2.50ch01p80659733	T/C		CC/C:T, no TT		
SL2.50ch01p89029434*	C/T				TT
SL2.50ch01p89029630	G/A				AA
SL2.50ch01p89244607*	T/C				CC
SL2.50ch01p89348895*	A/C/T				CC
SL2.50ch01p94038783*	T/G				TT
SL2.50ch02p24531020	G/A				AA
SL2.50ch02p24532087	C/A			CC	
SL2.50ch02p35539680*	A/G				GG

Marcadors genètics i fenotípics per a les tipologies de Tomata Valenciana

SNP	Allels	MASCLET	INTERMITJA	BLANCA	VAL-ROSA
SL2.50ch02p35607433*	C/T				TT
SL2.50ch02p3895068	C/T				TT
SL2.50ch02p44364097*	T/C				CC
SL2.50ch02p45997204*	A/G				GG
SL2.50ch02p53208872*	C/T				TT
SL2.50ch03p1061333*	A/G				GG
SL2.50ch03p119838*	G/C				CC
SL2.50ch03p3733445*	A/G		AA/A:G, no GG		
SL2.50ch03p4426965*	A/C				CC/A:C, no
SL2.50ch03p4426971*	G/A				AA/A:G, no
SL2.50ch03p4937507*	G/C				CC
SL2.50ch03p5231807*	G/A		AA/A:G, no GG		
SL2.50ch03p59609868*	A/G				GG
SL2.50ch03p59638871*	G/C				CC
SL2.50ch03p66720592*	A/G		AA/GG, no A:G		
SL2.50ch03p66750347*	C/T		TT/T:C,no CC		
SL2.50ch03p66753107*	C/T		TT/T:C,no CC		
SL2.50ch03p66808486	A/G		GG/A:G, no AA		
SL2.50ch03p66833545	T/A		CC/TT, no C:T		
SL2.50ch03p66874111	C/T		TT/T:C,no CC		
SL2.50ch03p69011699	C/G				GG
SL2.50ch03p69555762*	A/C				CC
SL2.50ch03p69556446	G/A				AA
SL2.50ch03p69579481*	T/C				CC
SL2.50ch04p1123541	C/T				TT
SL2.50ch04p26802931	C/A			CC/C:A, no	
SL2.50ch04p55261858	T/A				AA
SL2.50ch04p55384842	C/T				TT
SL2.50ch04p55384880	G/A				AA
SL2.50ch04p55414200	T/A				AA
SL2.50ch04p59767485*	G/T				TT
SL2.50ch04p59828265*	C/T				TT/T:C, no CC
SL2.50ch04p61476938	A/C				CC
SL2.50ch04p63497531*	A/G				GG/G:A, no
SL2.50ch04p807231	T/C				CC/T:C, no TT
SL2.50ch04p807245	A/G				GG/G:A, no
SL2.50ch04p96325*	C/G				GG
SL2.50ch05p34298741*	G/A				AA
SL2.50ch05p52667398*	A/T				TT
SL2.50ch05p9223528*	A/C				CC
SL2.50ch06p31920713*	G/C		GG/G:C, no CC		
SL2.50ch06p45238362*	G/A			GG/G:A, no	
SL2.50ch06p45283958*	T/C			TT/T:C, no	
SL2.50ch07p12927481	C/A			CC/C:A, no	
SL2.50ch07p49977577	T/A			TT/T:A, no	
SL2.50ch07p49977641*	A/T			AA/T:A, no	

SNP	Allels	MASCLET	INTERMITJA	BLANCA	VAL-ROSA
SL2.50ch07p59961385	G/A			GG/G:A, no	
SL2.50ch07p62392771*	A/G				GG
SL2.50ch07p65831339*	A/C				CC
SL2.50ch07p66700403	C/T				TT
SL2.50ch07p66971471*	A/G			AA/G:A, no	GG
SL2.50ch08p1247310	C/A		CC/A:C, no AA		
SL2.50ch08p25163609	C/T				TT/T:C, no CC
SL2.50ch08p3352366*	A/G	AA/A:G, no	AA/A:G, no GG		
SL2.50ch08p3361135*	T/C		TT/T:C, no CC		
SL2.50ch08p3496896*	C/T		CC/C:T, no TT		
SL2.50ch08p64131809*	A/G				GG
SL2.50ch08p64190377*	T/C				CC
SL2.50ch08p738340*	C/T		TT/T:C, no CC		
SL2.50ch08p7712566	A/G				AA
SL2.50ch09p153009*	T/G				GG
SL2.50ch09p1748562*	C/T				TT
SL2.50ch09p27862117*	A/G				GG
SL2.50ch09p416051*	T/A				AA
SL2.50ch09p453674*	C/A				AA
SL2.50ch09p53951025*	G/T				TT/T:G, no
SL2.50ch09p54697147*	T/C				CC
SL2.50ch09p58822621*	A/C				CC
SL2.50ch09p60135040	C/T		CC		
SL2.50ch09p60135063	A/C		AA		
SL2.50ch09p62311542	C/A				CC
SL2.50ch09p70416	G/A				AA
SL2.50ch09p71893565*	G/A	AA/A:G, no	AA/A:G, no GG		
SL2.50ch09p71964878*	A/G		AA/A:G, no GG		
SL2.50ch09p72211	G/A				AA
SL2.50ch09p755632	C/A				AA
SL2.50ch09p755688	T/C				CC
SL2.50ch09p755697	T/G				GG
SL2.50ch09p755706	G/A				AA
SL2.50ch09p755720	A/C				CC
SL2.50ch09p755740	A/G				GG
SL2.50ch10p28146400	C/G	CC			
SL2.50ch10p43862198	A/C				CC
SL2.50ch10p552726	C/T				TT
SL2.50ch10p554338	C/A				AA
SL2.50ch10p57440432*	G/A				AA
SL2.50ch10p59183897*	T/C				CC
SL2.50ch10p59403452*	C/T				TT
SL2.50ch10p63394901*	C/A		AA/A:C, no CC		
SL2.50ch10p8344199	T/A				AA/T:A, no
SL2.50ch11p1231134	C/A		CC/A:C, no AA		
SL2.50ch11p18628870*	A/G				AA

Marcadors genètics i fenotípics per a les tipologies de Tomata Valenciana

SNP	Allels	MASCLET	INTERMITJA	BLANCA	VAL-ROSA
SL2.50ch11p18628881	A/C				AA
SL2.50ch11p18628946	T/C				TT
SL2.50ch11p18628951	C/A				CC
SL2.50ch11p2070298*	T/A				AA
SL2.50ch11p4775556*	A/G				GG
SL2.50ch11p4922873*	G/C				CC
SL2.50ch11p54253430*	G/A		AA/A:G, no GG		
SL2.50ch11p7703552*	T/C				CC
SL2.50ch11p7868972	A/G				GC
SL2.50ch12p15990803	G/A				AA
SL2.50ch12p17321168	G/T				TT
SL2.50ch12p17413947	A/G				GG
SL2.50ch12p17423000	A/C				CC
SL2.50ch12p19833640	T/C				CC
SL2.50ch12p20990437	G/C				CC
SL2.50ch12p21960763*	G/C		GG/G:C, no CC		
SL2.50ch12p22067353	T/G				G
SL2.50ch12p24117670	A/G				G
SL2.50ch12p24407863	C/A				A/A:C, no C
SL2.50ch12p25668410	C/G				G
SL2.50ch12p35043080	G/C				C
SL2.50ch12p35829120	C/A				A
SL2.50ch12p36251012	C/T				T
SL2.50ch12p49894516	C/A				A/A:C, no C
SL2.50ch12p52196229	G/C				C
SL2.50ch12p52875840	C/G				G/G:C, no C
SL2.50ch12p54056770	A/C				C
SL2.50ch12p57110930	C/A				A
SL2.50ch12p60521516	G/A				G
SL2.50ch12p60521533	G/A				G
SL2.50ch12p60564649	G/A				G
SL2.50ch12p60581307	A/G				A
SL2.50ch12p60581391	A/G				G/G:A, no A
SL2.50ch12p60581584	G/T				T/G, no G:T
SL2.50ch12p60641869	G/T				T/G, no G:T
SL2.50ch12p60641902*	A/C				A/C, no A:C
SL2.50ch12p60641997*	C/T				T/C, no C:T
SL2.50ch12p60688529*	C/T				T
SL2.50ch12p60688552*	G/C				G
SL2.50ch12p61403796	T/G				G
SL2.50ch12p62484851*	T/C				C
SL2.50ch12p66814292*	A/T				T/T:A, no A
SL2.50ch12p7888415	G/T				T
SL2.50ch12p7933013	A/G				G
SL2.50ch12p8663010	G/A				A

En resum, en aquest treball utilitzant les dades de l'Atles de tomata de data, hem classificat un grup de tomates 'Valencianes' en base al fenotip i al genotip, i hem identificat que els trets mesurats de manera objectiva amb aparells específics són aquells que millor diferencien les diferents tipologies 'Valencianes'. Concretament trets relacionats amb la mida del fruit, l'índex de forma i la forma proximal i distal del fruit. A més, hem identificat signatures moleculars i combinacions de loci i al·lels que diferencien les quatre tipologies, generant una sèrie de marcadors per poder classificar noves entrades.

4. Referències bibliogràfiques

- Ballester, A.-R., Molthoff, J., de Vos, R., Hekkert, B. te L., Orzaez, D., Fernández-Moreno, J.-P., ... Bovy, A. (2010). Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SIMYB12 Leads to Pink Tomato Fruit Color. *Plant Physiology*, 152(1), 71 LP – 84. <https://doi.org/10.1104/pp.109.147322>
- Blanca, J., Pons, C., Montero-Pau, J., Sanchez-Matarredona, D., Ziarsolo, P., Fontanet, L., ... Granell, A. (2022). European traditional tomatoes galore: a result of farmers' selection of a few diversity-rich loci. *Journal of Experimental Botany*, 73(11), 3431–3445. <https://doi.org/10.1093/jxb/erac072>
- Casañas, F., Simó, J., Casals, J., & Prohens, J. (2017). Toward an Evolved Concept of Landrace . *Frontiers in Plant Science* , Vol. 8, p. 145. Retrieved from <https://www.frontiersin.org/article/10.3389/fpls.2017.00145>
- Cebolla-Cornejo, J., Roselló, S., & Nuez, F. (2013). Phenotypic and genetic diversity of Spanish tomato landraces. *Scientia Horticulturae*, 162, 150–164. <https://doi.org/https://doi.org/10.1016/j.scienta.2013.07.044>
- Figàs, M. R., Prohens, J., Casanova, C., Fernández-de-Córdova, P., & Soler, S. (2018). Variation of morphological descriptors for the evaluation of tomato germplasm and their stability across different growing conditions. *Scientia Horticulturae*, 238, 107–115. <https://doi.org/https://doi.org/10.1016/j.scienta.2018.04.039>
- Gentilcore, D. (2009). Taste and the tomato in Italy: a transatlantic history. *Food and History*, 7(1), 125–139. <https://doi.org/10.1484/J.FOOD.1.100639>
- Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., & van der Knaap, E. (2013). Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Molecular Genetics and Genomics*, 288(3), 111–129. <https://doi.org/10.1007/s00438-013-0733-0>
- Husson, F., Josse, J., & Pages, J. (2010). Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data. *Applied Mathematics Department*, 1–17.

- Kassambara, A. (2017). Multivariate Analysis II: Practical Guide to Principal Component Methods in R. *Sthda*, 170.
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software; Vol 1, Issue 1 (2008)* . Retrieved from <https://www.jstatsoft.org/v025/i01>
- López-Terrada, M. (2017). The History of the Arrival of the Tomato in Europe: An Initial Overview. Retrieved from <http://tradiotom.eu/project/history/>
- Nitsch, L., Kohlen, W., Oplaat, C., Charnikhova, T., Cristescu, S., Michieli, P., ... Rieu, I. (2012). ABA-deficiency results in reduced plant and fruit size in tomato. *Journal of Plant Physiology*, 169(9), 878–883. <https://doi.org/https://doi.org/10.1016/j.jplph.2012.02.004>
- Pons, C., Casals, J., Brower, M., Sacco, A., Riccini, A., Hendrickx, P., ... Granell, A. (2023). Diversity and genetic architecture of agro-morphological traits in a core collection of European traditional tomato. *Journal of Experimental Botany*, erad306. <https://doi.org/10.1093/jxb/erad306>
- Pons, C., Casals, J., Palombieri, S., Fontanet, L., Riccini, A., Rambla, J. L., ... Granell, A. (2022). Atlas of phenotypic, genotypic and geographical diversity present in the European traditional tomato. *Horticulture Research*, 9, uhac112. <https://doi.org/10.1093/hr/uhac112>
- Sagi, M., Scazzocchio, C., & Fluhr, R. (2002). The absence of molybdenum cofactor sulfuration is the primary cause of the flacca phenotype in tomato plants. *The Plant Journal*, 31(3), 305–317. <https://doi.org/https://doi.org/10.1046/j.1365-313X.2002.01363.x>
- Soler, S. (2024). Conservando el germoplasma del tomate valenciano. *L'Agrària*, #05.